Psychology of music, §II: Perception & cognition

2. Rhythm.

RICHARD PARNCUTT, CAROLYN DRAKE

The perception of Rhythm involves the perceptual and cognitive organization of events in time, whereby each sound event is situated in relation to those that have already occurred (memory) and those yet to come (expectancy). Different cognitive processes occur over short and long time-spans.

(i) Surface organization.

The acoustical signal is first perceptually segmented into separate events corresponding to the attack points of musical elements such as tones and chords (Köhlmann, 1984; Vos and Rasch, 1981). The moment at which an event is perceived to occur is its ‘perceived onset’ (related to the perceptual centre of phonemes). The time interval between the onset of one event and the onset of its successor is called the ‘inter-onset interval’ (IOI). The physical duration of an event (i.e. the time interval between its onset and offset) may be shorter than its IOI (e.g. in staccato) or longer (overlapping legato). Rhythmic organization is generally influenced more by IOI than by physical duration (Vos, 1976–7; Vos, Mates and van Kruysbergen, 1995).

IOIs are often perceived categorically in relation to surrounding IOIs (Schulze, 1989). The categories tend to correspond to the note values of music notation and are usually unaffected by typical deviations from metronomic
timing (such as rubato). The category to which an IOI is allocated depends on its metrical context (Clarke, ‘Categorical’, 1987) and the categorization process may be modelled using neural networks (Desain and Honing, 1989). A listener may assign all notes in a rhythm to as few as two IOI categories (e.g. quavers and semiquavers or simply long and short). This is an appropriate strategy, given that 80% of the notes of typical short classical pieces or movements correspond to just two note values, in the ratio 1:2 or (less often) 1:3 (Fraisse, 1982).

(ii) Grouping and metre.

The events of a rhythm are hierarchically organized in two distinct ways, known as grouping and metre (Cooper and Meyer, 1960; Deutsch, 1982; Lerdahl and Jackendoff, 1983; Handel, 1998; Drake, 1998–9). From a perceptual viewpoint, rhythm is characterized by, and may even be defined as, a combination of these two forms of organization.

A rhythmic or temporal group is defined as a series of events that are close to each other in time. Perceptual groups are formed by segmenting the musical surface at events with relatively long IOIs, or at changes of timbre, register, loudness or articulation (Handel, 1981; Deliège, 1986–7; Palmer and Krumhansl, 1987; Clarke and Krumhansl, 1989–90). When grouping occurs on several hierarchical levels at once, the resultant organization is called a ‘hierarchical grouping structure’. At the musical surface, groups correspond to short motifs. Motifs combine to form phrases, which in turn group into longer phrases, extended passages, movements and eventually whole pieces. In experiments to investigate grouping structure, listeners may be asked to listen to a long piece of music and indicate where sections begin and end. Segmentations between groupings spanning longer time periods tend to be associated with longer pauses or
striking changes in physical event characteristics (Deliège and El-Ahmadi, 1990). Further evidence for the psychological reality of temporal groups has been obtained by adding clicks to a melody and asking listeners to recall their positions (the clicks tend to ‘migrate’ in the direction of group boundaries: Sloboda and Gregory, 1980), and by counting errors in musical performances, which tend to occur more often at group boundaries than within groups (Palmer and van de Sande, 1995).

Metre is a form of perceptual organization based on temporal regularity (underlying beat or pulse). A sensation of pulse may be evoked by temporal regularity at any level within a sound sequence, or whenever relatively salient events (or motivic patterns) are perceived as roughly equally spaced in time. The musical behaviour that perhaps most clearly reflects the perception of pulse is foot-tapping to music. Cognitively, the process of regularity extraction may be regarded as one of synchronizing an internal time-keeper or clock to music (Wing and Kristofferson, 1973; Povel and Essens, 1984–5; Essens, 1995), tolerating musically typical deviations from periodicity (Shaffer, Clarke and Todd, 1985; Large and Jones, 1998). Temporal regularity may be perceived in the face of considerable deviations from mechanical regularity or rubato. If a sequence abruptly stops, the listener expects the pulse to continue; attention is enhanced at the temporal locations of expected events (Jones and Boltz, 1989).

The perceptual salience of a pulse sensation depends on its Tempo. Musical pulses are confined to a tempo range of roughly 30 to 300 beats per minute, or an IOI range of 200 milliseconds to 2 seconds (Fraisse, 1982), known as an ‘existence region’ (Jones and Boltz, 1989). The most salient pulses usually have tempos in the vicinity of ‘spontaneous tempo’ (the tempo at which a participant in an experiment will tap if asked to do so at equally spaced...
intervals that are otherwise unspecified: Fraisse, 1957). Spontaneous tempo varies widely from one person to another: inter-tap intervals mostly lie in the range of 400 to 900 milliseconds, with a mean (relative to a logarithmic scale) of about 600. A similar range is observed when listeners are asked to tap in time with a piece of music. Sensitivity to small changes in tempo is most acute in the range 300 to 800 milliseconds (Fraisse, 1967; Drake and Botte, 1993). These phenomena appear to have their origins in the physical properties of the human body and suggest a strong connection between perception of rhythm and human movement (walking, dancing, heartbeats): Truslit, 1938; Gabrielsson, 1973; Fraisse, 1974; Clynes and Nettheim, 1982; Kronman and Sundberg, 1987; Todd, 1992; Davidson, 1993; Shove and Repp, 1995; Krumhansl and Schenk, 1997; Parncutt, 1997.

A metrical structure consists of hierarchical levels of pulsation or rhythmic strata (Yeston, 1976). For example, the cognitive structure corresponding to 3/4 metre includes pulses of crotchets and dotted minims, and usually also includes faster pulses (e.g. quavers) and slower pulses (e.g. groups of two bars, or hypermetre: Rothstein, 1989). The multiple pulses that make up a conventional musical metre are mutually consonant in the sense that every event at every level (except the fastest) corresponds to an event at the next-faster level. Simultaneous pulses can also be dissonant (Hlawicka, 1958; Krebs, 1987). From least to most dissonant, three cases can be distinguished: rhythmic displacement, as in fourth-species counterpoint (same period, different phase); polyrhythm (cross-rhythm) such as three against two (same phase, different period: Handel and Oshinsky, 1981; Beauvillain and Fraisse, 1983–4); and both displacement and polyrhythm (different period, different phase), such as the start of Gershwin’s ‘I got rhythm’ (a displaced series of semiquavers against an accompaniment of crotchets). Complex metres such as
9/8 (when arranged (2+2+2+3)/8), in which the crotchet pulse is effectively displaced by a quaver at every bar-line, have not yet been the subject of psychological investigation (but see London, 1995–6).

Whenever temporal regularity is perceived at different levels – whether consonant or dissonant – listeners tend to focus on, or attend to, a single level of moderate tempo (period near 600 milliseconds) and perceive other levels (and hence all events) relative to that level. In the case of the consonant levels that make up a metre, this level is called the ‘tactus’ (Lerdahl and Jackendoff, 1983) or ‘referent level’ (Jones and Boltz, 1989). It may be determined experimentally by asking listeners to tap at regular intervals in time with the music. Listeners can switch their attention to faster and slower rhythmic levels at will. In an oscillator model of metre perception, a primary oscillator (corresponding to the tactus) is situated within the optimum tempo range, and may become coupled with other oscillators tuned to other hierarchical levels (Large and Jones, 1998).

Metre also involves characteristic alternations of weaker and stronger beats within each bar or period. Cognitive representations of metres such as 2/4 and 6/8 have been established experimentally using ‘probe-tone’ methodology (more usually used in investigations of tonality see C.L. Krumhansl: Cognitive Foundations of Musical Pitch (London, 1990)); the relative strengths of beats within the metre are quantified on a continuous scale (Palmer and Krumhansl, 1990). Such patterns are presumably learnt by repeated exposure to music in given metres and subsequently recognized during listening.

Grouping and metrical structures are intertwined at all levels of rhythmic organization. For example, the first ten notes of the main melody of Mozart’s G minor Symphony k550 (ex.1) imply as many as four different hierarchical
levels of grouping (from two-note phrases to all ten notes) and five metrical levels (quavers, crotchets, minims, semibreves, double-semibreves). The second-last note (d") bears the greatest metrical accent because it belongs to all five metrical levels.

Over longer time spans, perceptual hierarchies of grouping and metre are generally neither clearcut nor complete (Clarke, 1988). For example, a listener may be uncertain as to whether bars 1 and 2 of a piece, or bars 2 and 3, group together to form a hypermetre. Similarly, there may be ambiguity as to which motifs at the musical surface belong to which phrases at adjacent structural levels. At any given moment in a piece, a listener will have organized past events into incomplete, tentative hierarchies, and on this basis will have expectations regarding how these structures will be maintained as the piece progresses. In the case of two competing, incompatible metrical descriptions of the same musical surface, a listener may switch from one hierarchical description to another, when evidence in favour of the second becomes stronger than that in favour of the first. It is thus not generally possible to give a definitive hierarchical description of the rhythmic perception of a piece of music.

Metrical ambiguity may be said to occur when two incompatible metrical interpretations exist for the same musical surface – in other words, there are two (dissonant) alternatives for the tactus. Metrical ambiguity is more commonplace in musical works than their notation would suggest (Vos, Collard and Leeuwenberg, 1981; Parncutt, 1993–4). In ambiguous cases, listeners tend to choose one interpretation soon after the piece begins and stick to it in the face of evidence to the contrary (Longuet-Higgins and Lee, 1982; Lee, 1991). The cognitive process of switching attention between dissonant rhythmic levels requires either considerable mental effort or a
considerable change in performed accentuation (Tuller and others, 1994).

(iii) Accent.

Everyday usage equates Accent with loudness: attention can be attracted to an event simply by playing it more loudly (or sometimes more softly) than events in its context. Here accent will be considered synonymous with event salience. Anything that makes an event sound more important than adjacent events, or which attracts the attention of a listener to an event, may be regarded as an accent (Jones, 1987).

The grouping and metrical structures perceived in a piece of music depend ultimately on the timing and ‘phenomenal accent’ of the events at the surface (Lerdahl and Jackendoff, 1983). The most important contributor to phenomenal accent is typically the IOI between the event and its successor (Steedman, 1977): the longer the IOI following an event, the stronger the accent. The IOI preceding an event can also contribute to its perceived accentuation, but to a lesser extent (Lee, 1991). Apart from IOI, phenomenal accents are generated by relative loudness (dynamic accents); by articulation (e.g. by switching from legato to staccato); by timbral variation (manipulating the temporal or spectral envelope of events); by adjusting intonation; by melodic contour (melodic accents occur at peaks and valleys in the melodic contour and follow melodic leaps: Thomassen, 1982; Huron and Royal, 1995–6); and by harmonic progressions (harmonic accents occur at dissonances and harmonic changes: Smith and Cuddy, 1989; Dawe, Platt and Racine, 1994–5).

Structural and metrical accents are associated with grouping and metrical structures respectively. At the simplest level, a structural accent occurs at the start and
at the end of every rhythmic group (Povel and Essens, 1984–5; Drake, Dowling and Palmer, 1990–91), and a metrical accent occurs at every event in a pulse (or potential tactus). Structural and metrical accents are most likely to be perceived if they occur simultaneously on several hierarchical levels: the greater the number of levels, or the greater the accent at each level, the more salient will be the accent (Todd, 1985–6; Parncutt, 1987; Rosenthal, 1992).

Accents may be either immanent to a (notated) musical work or added to the music during performance. Structural accents are normally regarded as immanent, although they can also be affected by performance (Lester, 1995). Apart from dynamic (loudness) accents, the most important performed accents are Agogic (Riemann, 1884). Agogic accents are produced by delaying event onsets or lengthening IOIs relative to the prevailing metrical framework (Gabrielsson, 1974; Sloboda, 1983; Clarke, 1988; Palmer, 1989; Repp, 1990;).

Timing variations in rhythmic performance have various functions. A performance that sounds perfectly regular (mechanical, metronomic) is not generally physically regular (Seashore, 1938; Drake, 1993; Gérard, Drake and Botte, 1993; Penel and Drake, 1998) but deviates from physical regularity in the same direction as, but to a smaller degree than, a typical expressive performance (Repp, 1997–8). Thus one function of timing variations is to make a performance sound regular – paradoxically, by making it physically irregular. Timing variations also have the function of clarifying the grouping and metrical structures, rendering them less ambiguous (Sundberg, 1988; Drake, 1993). Agogic accents can tell the listener where to hear the downbeat of a bar (Sloboda, 1983) or the start of a long phrase (Todd, 1985–6). Finally, timing variations affect the emotional character of a rhythm (Gabrielsson and Juslin, 1996). Timing variations are
perceptible when they exceed about 20 milliseconds in typical musical performances (Clarke, 1989), falling to six in monotonic isochronous sequences faster than four events per second (Friberg and Sundberg, 1999).

The ease with which a rhythm can be cognitively processed depends on the way different kinds of accent are distributed within the rhythm. Performances tend to be easier to understand, remember and reproduce when performed accents correspond to immanent accents (Drake, Dowling and Palmer, 1990–91; Clarke, 1992–3; Tekman, 1996–7). In the absence of performed accents, rhythms are easier to process when different kinds of immanent accent (e.g. melodic, IOI) coincide (Jones, 1987).

(iv) Rhythmic organization and tempo.

The perceived organization of a piece of music depends on the tempo at which it is performed (Handel, 1993). Tempo may affect both grouping and metre. The metrical level at which the tactus is located depends on tempo (Handel and Oshinsky, 1981) because distributions of tapping rates to music (measured absolutely, in beats per second) are almost independent of tempo (Parncutt, 1993–4). For example, a listener might tap out quavers when a piece is played slowly and crotchets when the same piece is played twice as fast, thus keeping the tapping rate in the same absolute range. In the case of grouping, the number of elements in a group increases as tempo increases (Clarke, 1982), keeping their absolute length about constant.

Patterns of agogics and dynamics depend on a performer’s perceptual organization of a piece, and thus are also affected by tempo (see Michon, 1974). Effects of tempo on timing and dynamics have been studied (Monahan and Hirsh, 1990; Desain and Honing, 1994);
analogous effects of tempo on the perception of music performances were reported by Repp (1995–6).

(v) Rhythm v. form.

As rhythmic groups become longer and pulses slower, a perceptual transition occurs from the domain of rhythm to that of form (Clarke, ‘Levels’, 1987). In grouping, the change may be said to occur when a group’s duration exceeds that of the psychological present (Fraisse, 1957; Crowder, 1993), defined as a short period of time during which relationships between successive events can be perceived directly, without cognitive reference to earlier periods (memory, rehearsal; similar to Baddeley’s 1986 ‘working memory’). The duration of the psychological present depends on musical tempo and complexity, but it is normally estimated to lie in the range of two to seven seconds. In the case of pulse and metre, the transition from rhythm to form may be said to occur when temporal regularity ceases to imply physical movement or dance (beyond a period of about two seconds: Fraisse, 1974) and rhythmic temporal anticipation is no longer possible (Mates and others, 1994).

BIBLIOGRAPHY

H. Riemann: *Musikalische Dynamik und Agogik* (Hamburg, 1884)
C.E. Seashore: *Psychology of Music* (1938/R)
A. Truslit: *Gestaltung und Bewegung in der Musik* (Berlin, 1938) [incl. 3 sound discs]
P. Fraisse: ‘Le seuil différentiel de durée dans une suite régulière d'intervalles’, *Année psychologique*, lxvii (1967), 43–9


M. Yeston: *The Stratification of Musical Rhythm* (New Haven, CT, 1976)


