1987 b

A. Gabrielsson (Ed.),
Action and Perception in Rhythm and Music.
Publications issued by
the Royal Swedish Academy of Music No. 55,
1987.

THE PERCEPTION OF PULSE IN MUSICAL RHYTHM

Richard Parncutt

Department of Psychology,
University of New England,
Armidale NSW 2351,
Australia

Abstract

Rhythm may be regarded as a sequence of perceived *events,* each of which is specified by its position in time relative to other events and by its salience. A theory is outlined by which *pulse* and *metre* percepts in music are determined by (and therefore predictable from) configurations of event percepts. The theory is based on the assumption that the perception of pulse in music has a universal component, conditioned by periodically repeating sounds in the human environment. The most important sounds conditioning the perception of rhythm may be the sounds associated with the heartbeat and walking movements of a mother, as heard by her unborn child. Roughly equally-spaced events in music reminiscent of these sounds evoke emotional (heartbeat) and movement (motion during walking) responses. When suitably formulated, the theory may be used to predict the relative saliences of the various pulse and metre percepts evoked by a rhythmic sequence.

Event Percepts and their Salience

Each note or chord in a piece of music, and each syllable in speech, is associated with the perception of an "event". Events are characterized by their perceived onset times and perceived durations. The end of an event is assumed here to coincide with the start of the event which immediately follows it. So the duration of an event is the time interval between its own onset and the onset of the next event.

The relationship between the waveform of a passage of sound and the times at which events are perceived is quite complex (Köhlmann, 1984). Event onsets are precipitated by relatively fast or sudden changes in loudness, pitch or timbre, or by some combination of these. In order to predict event onset times with any generality, then, it is first necessary to predict loudness, pitch and timbre as a continuously varying function of time by means of appropriate algorithms.

In music it is important not only at what exact times event percepts begin but also how perceptually important or *salient* they are. Accented, stressed or rhythmically strong events are more salient than other events. In this paper, the term "event salience" refers to the probability that an event will be noticed, or some measure thereof.

The salience of an event depends on the way in which loudness, pitch and timbre are changing when the event occurs, relative to the way in which these attributes typically change in the context of the event. The salience of an event also depends on its duration. Other things being equal, longer events in music are more salient than shorter ones (Vos, 1977). The same is true in speech: accented syllables normally have longer durations than unaccented ones.

In an algorithm for the prediction of pulse salience (Parncutt, 1985), event salience was assumed to increase as (onset-to-onset) event duration increased for relatively short durations, and to saturate (approach a constant value) at high durations. Relative accentuation of different events (e.g. relative loudness) was not accounted for in the algorithm.

In musical rhythm, the salience of an event also depends on its placement relative to the prevailing metric scheme. Events falling on "strong beats" in music are reinforced, i.e. their salience is increased, by their rhythmic placement. Even a musical "rest" may be perceived as an event, if it falls on a strong beat. A rhythmically strong rest may be regarded as a "missing stimulus", whose perceptual reality is verified by neurophysiological measurements (McCallum, 1981).

The contribution of rhythmic placement to event saliences may be predicted quantitatively on the basis of the theory developed in this article (below). The simplest method might involve adding up the saliences of all the "pulse percepts" to which an event percept belongs.

Grouping of event percepts

Groups of event percepts in music may be perceived as single units. The successive grouping of groups of event percepts produces a kind of hierarchy by which a rhythmic pattern may be understood by the listener.

The same event sequence can produce many different perceptual hierarchies, because event percepts may be organized in many different ways. Three examples follow.

Sound events may be grouped (i) on the basis of their perceptual attributes: loudness, pitch, timbre, onset time and duration. Groups of event which are "close" to each other in some combination of these attributes form perceptual "streams". For example, sounds of similar pitch separated by small time intervals (Miller and Heise, 1950; Noorden, 1975) and sounds of similar timbre (Dannenbring and Bregman, 1976) may be grouped to form a stream.

Sound events may be grouped (ii) on the basis of known or previously experienced attributes of environmental sound sources. In music, for example, clarinet sounds may be grouped even though they differ markedly in pitch and timbre. Sounds perceived to come from about the same position in space may be grouped, e.g. in music recorded in stereo.

From the musical point of view, the most interesting aspect of the grouping of sound pecepts is (iii) the spontaneous recognition of familiar patterns of sound. The auditory system is remarkably good at this. For example, most people can quickly recognize a familiar melody from the pitch pattern alone, or – if the melody has a distinctive rhythm – from the rhythm alone.

An especially important kind of rhythmic pattern is the *pulse:* a simple sequence of equally-spaced event percepts. The experience accompanying the perception (or "recognition") of a pulse may be called a *pulse percept.* Pulse may be regarded as a basis of rhythm, in that it distinguishes rhythm from non-rhythm. Furthermore, it is found across musical cultures. The universality of pulse suggests that it has an extra-musical origin.

Origins of Pulse Perception

Some have suggested that the origins of musical pulse sensations lie with certain aspects of brain functioning, for example in periodic firing patterns of ensembles of neurons. In my opinion this is a strange and unworkable hypothesis. It is strange because the nature of our experiences is normally most easily relatable to the nature of the human environment and how we interact with it rather than to the structural characteristics or limitations of the brain. It is unworkable because our knowledge of brain function is at present very limited, as is our understanding of the relationship between brain processes and experiences. At any rate it is people, not brains, that experience musical rhythms (see Coulter, 1979).

A more common belief is that pulse perception is related to certain periodicities in bodily functions such as heartbeats and footsteps. This idea has a long history in experimental psychology (see Fraisse, 1982). It is based on the observation that periodicities of strongly felt pulses in music and of heartbeats and footsteps lie in about the same range: around 0.5 to 0.8 seconds.

It is not immediately clear how and why heartbeats and footsteps might contribute to the perceived structure of music. Heartbeats are seldom actually heard, and footsteps do not normally play a role in musical performance. The most likely possibility seems to be that the tendency to recognize and react to musical pulse is acquired by the auditory system by *prenatal conditioning:* familiarization of the foetus with sounds associated with the heartbeat and walking movements of the mother.

In the months immediately before birth, the auditory system of the infant is already in an advanced stage of development (Verny, 1981). The auditory system is the means by which the developing infant will interact with its post-natal auditory environment. The system is therefore inherently sensitive to stimulus invariances which suggest the presence of specific environmental objects.

As a rule, sounds such as the mother's heartbeat (cardiovascular system), digestive sounds, voice sounds and sounds associated with physical movement (e.g. walking) are audible to the human foetus, often having SPLs of 60 dB or more; while sounds originating outside the mother's body are attenuated by 20 to 30 dB (Bench, 1968; Vince et al., 1982b). Sound is transmitted through the mother's body to the developing ears of the foetus via the amniotic fluid. Sounds associated with the mother's walking are linked with physical movement of the foetus. Changes in heartbeat and voice patterns regularly accompany variations in the emotional state of the mother; the foetus might somehow share these emotions. Stomach sounds and sounds from outside the mother's body are probably too irregular to affect later perceptual processes in simple ways, or in ways relevant for the perception of rhythm.

Musical pulses range between two extremes, which may be labelled *strict* and *rubato* pulses. Strict pulses (in Italian: "tempo giusto") tend to invoke physical movement and dance, and are appropriate for styles such as jazz (e.g. "swing") and rock. Rubato pulses suggest physical movement less strongly; they are more suitable for emotional expression and are appropriate for romantic musical styles (e.g. Chopin Nocturnes).

There is no sharp dividing line between strict and rubato pulse. Most rhythmic music can be interpreted as containing elements of both. Even quite strict rhythms can be expressive; and quite flexible rhythms can imply movement. In spite of these qualifications, however, the distinction between "strict" and "rubato" rhythms is musicologically significant, and has been the subject of much discussion and debate (see Scholes, 1970, under "Rubato").

The theory that pulse perception has prenatal origins suggests an explanation for the distinction between strict and rubato rhythms. Body movements in walking and running tend to be very regular, but occur at different rates on different occasions. Associated sounds heard by the foetus would have similar characteristics, and would be linked with physical movement of the foetus. Perhaps this is why "strict" musical pulses of different speeds evoke physical movement and dance. Heartbeats (and associated sounds heard by the foetus) are not as steady as footsteps, but speed up and slow down with breathing, as well as with physical exertion and emotional changes. Musical rubato appears to depend on phrasing, implied activity and emotional content in a remarkably similar way.

Terhardt (1974) has suggested that the perception of harmony in music (in particular, of the roots of chords) is based on the spontaneous recognition of patterns of pitch corresponding to the pitches of audible harmonics of complex tones. The most important kinds of complex tone for humans are speech vowels. It seems likely that the learning stage for Terhardt's pattern recognition process is well under way, perhaps even complete, before birth, as a result of exposure of the foetus to the voice of the mother. Association with the mother's voice may therefore have something to do with the way in which musical harmonies express emotion.

Musical melodies, chords and rhythms do not normally remind people of voices, heartbeats and walking, as adults normally perceive these things. Instead, they manifest themselves in abstract musical forms. "Chord" and "pulse" sensations in music may be described as "auditory images" (McAdams, 1984) – not of environmental objects, as is usually the case, but of the mother as perceived by the foetus.

Music has many social and psychological functions. One of these may be described as "escapism". Music can allow musicians and audiences to temporarily forget about their real lives and become absorbed in an abstract but meaningful world of sound. Such behaviour may perhaps be understood in terms of the present theory of prenatal conditioning. Specific sounds familiar from prenatal experience have a calming effect on animals (Vince et al., 1982a) and – in the case of heartbeat sounds – on human infants (Salk, 1960). A subconscious association between musical experience and the safety and security of the womb may at least partly explain the beauty and power of music.

Clues about the origins of pulse perception in humans may be provided by studies in comparative psychology. Hulse (1984) found that European starlings were sensitive to pulse in that they could distinguish rhythmic from non-rhythmic patterns. He found that patterns of sound were heard by starlings to be "rhythmic" if individual sounds in the pattern had constant duration or if individual sounds recurred at constant time intervals. Perception of rhythm by birds may be conditioned by environmental sounds exhibiting just these characteristics, produced by other birds as well as by animals such as crickets and frogs. Birds are exposed repeatedly to such sounds, both before hatching (when sounds are heard through the shell) and in later life.

Salience of Pulse Percepts

A pulse is a chain of events, roughly equally spaced in time. A pulse may be completely specified by two pieces of information: its period and its phase. The period is the time interval between successive events. The phase may be specified by the actual time at which any of the events occurs, relative to some reference time.

Pulse percepts in music correspond to pulses actually present among the events of a rhythmic sequence. Like all perceptual patterns, a pulse may be perceived even when some elements of the pattern – the equally spaced events – are missing.

Examples of different pulses evoked by a single rhythmic sequence are shown in Figure 1. Pulse saliences listed in the figure were predicted by a simple algorithm (Parncutt, 1985). For reasons of space, the algorithm is not described in detail in this article. Instead, the main assumptions by which the algorithm was written are outlined in a qualitative way.

It is useful to imagine a pulse as a kind of template which is mapped against the pattern of events evoked by a rhythmic sequence. This heuristic is not to be confused with what happens physically when a pulse is recognized. In particular, what happens in the brain is of no concern here. Of primary interest is the psychophysical relationship between the input and the output of the pulse recognition process.

The algorithm predicts pulse saliences by applying the following "rule of thumb":

> Each pair of event percepts in short-term memory contributes to the salience of one and one only pulse percept, whose period and phase are determined by the temporal positions of the two events.

As an example of this rule, consider the case when all the events in a sequence have the same salience. The predicted salience of a particular pulse percept is proportional to the number of pairs of events in the sequence corresponding to consecutive events in the pulse percept. According to the rule, n equally salient event percepts spaced at equal time intervals t are predicted to evoke a pulse percept of period t which is n-1 times as salient as the pulse evoked by two events separated by time interval t. For example, the salience of the pulse percept evoked by a sequence of three equally-spaced, equally salient events

Figure 1. Pulse Analysis of the Theme of Fugue No. 1 from Bach's "Well-Tempered Clavier" Bk.1. Pulse saliences are predicted by the algorithm of Parncutt (1985). Assumed tempo: 44 quarter notes per minute. Left column: pulse salience rank. Right column: predicted pulse salience (arbitrary units). The algorithm does not account for any effect of pitch contour on pulse salience.

is predicted to fall by half if the first or last event is removed.

The above rule is valid only over a limited time period, metaphorically called the "psychological present", "short-term memory" or "echoic memory". The duration of short-term memory is normally of the order of a few to several seconds (Neisser, 1967; Glucksberg and Cowen, 1970; Crowder, 1970). The salience of the pulse percept evoked by a pulse of long duration does not increase without bound, but levels off to a finite value (saturates) within a few seconds, as its duration approaches that of short-term memory. Short-term memory also sets an upper limit to the period of a musical pulse percept: pulses whose periods exceed a few seconds cannot evoke a feeling of pulse, as no more than one of the events of such a pulse can be "stored" in a single "chunk" of short-term memory.

Pulse percepts normally have periods in the range 0.2 to 1.8 seconds, while the spontaneous tempi of finger taps (performed pulses) and the preferred tempi of heard pulses normally lie in the range 0.5 to 0.7 seconds (Fraisse, 1982, and references therein). According to the learning model hypothesised here, these limitations may be understood in terms of the periods of repeating sounds in the human environment, especially of heartbeats and footsteps. They are accounted for in the algorithm by means of a bell-shaped weighting function, centred on a pulse period of about 0.6 seconds and symmetrical in the logarithm of pulse period. The function is reminiscent of the "spectral frequency weight" function used by Terhardt, Stoll and Seewann (1982) to model the spectral dominance effect in pitch perception: it has a similar shape, and plays a similar role in the algorithm.

Perception of Metre

Sequences of auditory events normally evoke many different pulse percepts of different saliences at the same time (see Figure 1). In a first approximation, one might assume that a listener, presented with a given rhythmic sequence and asked to tap along with it at relaxed, equally-spaced intervals, would tap along with the most salient pulse percept evoked by the sequence. However, it is possible – indeed, normal – for a listener to attend to several pulses at once, noticing all of them while tapping out only one (usually a relatively slow one). In this case the listener is grouping pulse percepts into a more holistic percept which I call a *metre percept*

Metre is normally defined as it is notated in Western music: as a particular way of dividing up a measure into beats. In this paper, a different definition of metre is used. It is based on perception rather than notation. Metre is assumed to be perceived when two or more pulses are perceived at the same time.For example, in "perceived 3/4 time" the pulse at the beat (quarter notes) and that at the measure (three-quarter notes) are noticed simultaneously. Other pulses, such as a pulse of eighth notes (subbeats) or one-and-a-half notes (pairs of measures) may also be noticed at the same time as the beat and measure pulses.

The pulse percepts making up a metre percept must be *compatible.* Each event in the slower of two compatible pulse percepts coincides with an event in the faster pulse percept. In 3/4, for example, every third event in the "beat" pulse coincides with an event in the "measure" pulse. Compatible pulses are generally easy to attend to simultaneously, and therefore easy to perform. By contrast, incompatible pulses (or "cross-rhythms") can be quite difficult to attend to simultaneously (Handel, 1984) and to perform (Beauvillain and Fraisse, 1984).

An as yet unsolved problem in music perception is the objective prediction of the metre of an event sequence. Longuet-Higgins and Lee (1982) have successfully predicted the notated metres of some rhythmically quite ambiguous pieces, by means of an algorithm based on rules inferred from musical experience. In the process they did not, however, explicitly distinguish between notated and perceived rhythm. Vos (1985) showed experimentally that notated and perceived rhythm can be quite different. He measured what could be described as "metre saliences": probabilities that particular metres were perceived in given event sequences.

The present approach allows for the salience of a "metre percept" to be estimated simply by adding the saliences of the compatible pulse percepts which make it up. By this method, the most salient metre predicted to be evoked by the fugue subject in Figure 1 corresponds to its notated metre: it combines pulses number 6, 1, 2 and 9, and has a predicted salience of $0.15 + 0.47 + 0.23 + 0.11 = 0.96$. The second-highest predicted metre salience of the sequence is only marginally less than this: the predicted salience of the metre formed from pulses 6, 1, 3 and 10 is 0.93. This suggests that the metre of the sequence is quite ambiguous. Other possible metres are made up from pulses 4, 1 and 6 (predicted salience: 0.81) and from pulses 5, 1 and 6 (predicted salience: 0.77).

Temporal Shifts

Bengtsson and Gabrielsson (1983) have shown that timing in musical performance differs systematically from "mechanical" timing adhering exactly to notated time values. In other words, the time intervals between events in a rhythmic sequence need to be slightly "uneven" (in physical time) for the sequence to be perceived as musically "even" (i.e. musically "correct").

Sundberg and Frydén (1985) have formulated rules by which rhythmic time deviations may be introduced into synthesized musical performances. One of their rules is that fast sequences of notes should be played slightly too fast relative to a strict mechanical pulse. Another rule states that durational contrasts should be exaggerated by lengthening long notes and shortening short notes. This second rule works well in many cases, but it upsets the "swing" of alternating half and quarter notes in 3/4 time, shortening instead of lengthening the quarter notes (see Bengtsson and Gabrielsson, 1983).

An alternative approach to the understanding, prediction and synthesis of timing in music would be to allow instantaneous tempo to fluctuate relative to an average prevailing tempo. Bengtsson and Gabrielsson (1983) refer to such tempo fluctuations by the terms ”local tempo” and ”beat tempo”. Another possible term, ”instantaneous tempo”, covers fluctuations at all levels, including phrases, bars, beats and individual events. Instantaneous tempo may be regarded as a smooth function of time, i.e. as a continuous function with a continuous first derivative. Time intervals between events in music could be modelled by integrating the reciprocal of instantaneous tempo over time.

On a relatively large time scale, fluctuation of instantaneous tempo is called rubato. Rubato depends amongst other things on the rhythmic structure of a passage of music. Tempo fluctuations on a smaller time scale would also be expected to depend on rhythmic structure. Such fluctuations could account for many of the deviations from mechanical timing which have been observed experimentally.

Musical intuition and experience suggest that instantaneous tempo is slower near salient events than it is between events or near rhythmically weak events. Tempo seems to fall smoothly just before an important event, as a kind of preparation for the event, and to increase again afterwards. This happens without any sudden tempo changes, which would weaken the feeling of pulse. An appropriate algorithmic formulation of these principles could account for (i) the physical lengthening of the last quarter note in a 3/4 measure relative to the first half note of the next measure (Bengtsson and Gabrielsson, 1983); (ii) the physical speeding up of sequences of short notes in a context of long notes, but slowing elsewhere in order to maintain a constant overall tempo (Sundberg and Frydén, 1985); and (iii) the lengthening of pauses between musical phrases (Sundberg and Frydén, 1985).

Future Applications

The algorithm for the prediction of event, pulse and metre saliences outlined here (Parncutt, 1985) could be tested and improved by comparing its predictions with the results of experiments, as follows. Predicted event saliences could be tested by measuring the force with which each event in a rhythmic sequence is tapped out. Predicted pulse saliences could be tested by asking listeners (especially nonmusicians) to compare irregular rhythmic patterns with evenly-spaced rhythmic patterns (pulses), and to judge their similarity or ”relatedness”; or to listen to irregular rhythmic sequences, and tap out an underlying, regular ”beat”. Predicted metre saliences could be tested by asking musicians to notate unfamiliar rhythmic patterns.

A suitably simple, accurate and efficient algorithm could then be applied in the composition, performance and analysis of music. Theoretical insights gained during the creation of the algorithm could be applied in music theory and music education.

Acknowledgements

The concept of rhythm perception outlined here was inspired by Ernst Terhardt's theory of pitch perception: the relationship proposed here between a "pulse percept" and "event percepts" is remarkably analogous to the relationship between a "virtual pitch" and "spectral pitches". I would like to thank William Noble, Catherine Ellis and – especially – Alf Gabrielsson for their valuable suggestions and criticism. Financial support was provided by the Australian Commonwealth Department of Education.

References

Beauvillain, C. & Fraisse, P. On the temporal control of polyrhythmic performance. *Music Perception,* 1984, *1,* 485–489.

Bench, J. Sound transmission to the human foetus through the maternal abdominal wall. *Journal of Genetical Psychology,* 1968, *111,* 85–87.

Bengtsson, I. & Gabrielsson, A. Analysis and synthesis of musical rhythm. In J. Sundberg (Ed.) *Studies of Music Performance.* Publications issued by Royal Swedish Academy of Music, No. 39, Stockholm 1983.

Coulter, J. The brain as agent. *Human studies,* 1979, *2,* 335–348.

Crowder, R.G. The role of one's own voice in immediate memory. *Cognitive Psychology,* 1970, *1,* 157–178.

Dannenbring, G.L. & Bregman, A.S. Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, *2,* 544–555.

Fraisse, P. Rhythm and Tempo. In Deutsch, D. (Ed.): *The Psychology of Music.* New York: Academic Press, 1982.

Glucksberg, S. & Cowen, G.N.Jr. Memory for nonattended auditory material. *Cognitive Psychology, 1,* 149–156.

Handel, S. Using polyrhythm to study rhythm. *Music Perception,* 1984, *1,* 465–484.

Hulse, S.H. Rhythm discrimination and transposition in birds. *Music Perception,* 1984, *1,* 442–464.

Köhlmann, M. Rhythmische Segmentierung von Sprach- und Musiksignalen und ihre Nachbildung mit einem Functionsschema. *Acustica,* 1984, *56,* 193–203.

Longuet-Higgins, H.C. & Lee, C.S. The perception of musical rhythms. *Perception,* 1982, *11,* 115–128.

McAdams, S. The auditory image: A metaphor for musical and psychological research on auditory organization. In: Crozier, W.R. and Chapman, A.J. (Eds.): *Cognitive Processes in the Perception of Art.* North-Holland: Elsevier, 1984.

McCallum, W.C. Brain slow potentials elicited by missing stimuli. *Biological Psychology,* 1980, *11,* 7–19.

Miller, G.A. & Heise, G.A. The trill threshold. *Journal of the Acoustical Society of America,* 1950, *22,* 637–638.

Neisser, U. *Cognitive Psychology.* New York: Meredith, 1967.

Noorden, L. van. *Temporal Coherence in the Perception of Tone Sequences.* Eindhoven, The Netherlands: Institute for Perception Research, 1975.

Parncutt, R.M. Pulse and pulse salience in musical rhythm. Poster presentation at the *Third International Conference on Event Perception and Action,* Uppsala, Sweden, 1985.

Salk, L. Effects of normal heart beat sound on the behaviour of the new-born infant: implications for mental health. *World Mental Health,* 1960, *12,* 168–175.

Scholes, P. *The Oxford Companion to Music* (10th edition). London: Oxford University Press, 1970.

Sundberg, J. & Frydén, L. Teaching a computer to play melodies musically. In *Analytica – Studies in the Description and Analysis of Music in Honour of Ingmar Bengtsson 2 March 1985.* Publications issued by the Royal Swedish Academy of Music No. 47, Stockholm 1985.

Terhardt, E. Pitch, consonance and harmony. *Journal of the Acoustical Society of America,* 1974, *55,* 1061–1069.

Terhardt, E., Stoll, G. & Seewann, M. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America,* 1982, *71,* 679–688.

Verny, T. *The Secret Life of the Unborn Child.* New York: Delta, 1981.

Vince, M.A., Armitage, S.E., Walser, E.S. & Reader, M. Postnatal consequences of prenatal sound stimulation in the sheep. *Behaviour,* 1982, *81,* 128–139. (a)

Vince, M.A., Armitage, S.E., Baldwin, B.A., Toner, J. & Moore, B.C.J. The sound environment of the foetal sheep. *Behaviour,* 1982, *81,* 296–315. (b)

Vos, P.G. Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics,* 1977, *1,* 183–199.

Vos, P.G. Perceiving metric schemes in music. Unpublished manuscript, 1985.