MODELING IMMANENT DURATIONAL ACCENT IN MUSICAL RHYTHM

Richard Parncutt
Institute for Musicology, University of Graz, Austria

ABSTRACT

The accents produced by the time intervals (IOIs) that precede and follow musical events can be either *immanent* (evident from the score) or *performed* (expressive, *agogic*), and either *predurational* or *postdurational*. Agogic accents involve lengthening and delaying of events relative to temporal expectations established by metrical frameworks. Predurational immanent accents precede longer IOIs and mark downbeats (e.g., a half-note preceded and followed by quarter-notes). Postdurational immanent accents follow (even) longer IOIs and mark the beginnings of temporal groups (phrases). Predurational accents predominate at fast tempi and within phrases, postdurational at slow tempi and between phrases.

Predurational accent appears to be associated with *echoic memory*, a buffer for individual sound events such as speech syllables that operates independently of conscious attention and has a duration of around 1 second. Postdurational accent appears to be associated with the *psychological present* or *working memory*, which enables the comprehension of a linguistic phrase; its duration (a few to several seconds) is determined by limitations of conscious attention.

In a quantitative model, these two accent types combine to produce an overall estimate of durational accent, and further combine with phenomenal accentuation due to changes of loudness or timbre. The resultant temporal pattern of accents is then input to a pattern recognition algorithm that is confined to the psychological present (plus hysteresis) and predicts the underlying pulses and their saliences.

1. INTRODUCTION

Recent research on meter perception has focussed on *periodicity* and possible psychological and physiological bases of the experience of *pulse*. The issue of the *phase* of a perceived pulse (the position within each period at which the downbeat is perceived) has attracted less attention. This is surprising considering the strong influence of pulse phase on music experience.

The pulse of a piece of music is generally ambiguous with respect to both period and phase. Ambiguity of period is commonplace: for example, we can beat duple-time music either once or twice per measure, and attend to both pulses at the same time. This leads to a psychological definition of meter as a combination of compatible, simultaneously perceived pulses (Parncutt, 1987).

Phase ambiguity is more striking. Consider a piece of music in that begins with the instruments entering gradually. Each instrument plays a repetitive rhythmic figure. The listener finds an underlying pulse in the first few tones (Longuet-Higgins & Lee, 1982). If the opening patterns are deliberately syncopated, the listener may latch onto a pulse that is phase-shifted relative to the pulse intended by the composer or arranger and perceived by the musicians. As the other instruments join in and evidence in favor of a different downbeat accumulates, the listener is reluctant to shift the cognitive metrical framework. *Hysteresis, or* the "persistence of a structural representation despite stimulus parameters that would normally favor an alternative" (Large, 2001, p. 176), is particularly strong when the new pulse is *dissonant* (Krebs, 1987; Roeder, 2001) or incompatible with the old. Only when the evidence in favor of a different downbeat becomes overwhelming does the listener switch mentally to the "intended" pulse. After that, it is no longer possible to imagine the music relative to the original, shifted pulse: we seem to be hearing a different piece of music. When Sloboda (1983) asked pianists to sight-read various pieces of music, none of them noticed that two of the pieces were identical except that the barlines had been shifted.

The period of the perceived pulse of a piece of music is determined mainly by the *pulse-period dominance region* (Parncutt, 1994). Perceived pulses tend toward a tempo of about 100 events per minute, i.e. to a period of 600 ms. Their period is surprisingly independent of tempo (faster music does not necessarily evoke faster pulses). The surface also determines the pulse's *salience* (patterns that more closely resemble or include exactly periodic onset patterns generate more salient pulses).

Another factor contributing to the period and salience of pulse sensations is the *repetition of temporal patterns* (e.g., Handel, 1973; Lee, 1991; Lerdahl & Jackendoff, 1983; Steedman, 1977), which can be automatically detected by autocorrelation (Brown, 1993; Goto, 2001). Repetitive patterns can also involve pitch (repeating melodic fragments). In an approach guided by the principle of parsimony, such repetition can be regarded as an artifact of higher levels of musical structure: although rhythmic patterns tend to repeat with a periodicity corresponding to the period of the perceived pulse, this repetition may not represent the *original cause* of the pulse sensation and associated implied movements. Moreover, repetition cannot predict the *phase* of a pulse sensation (Temperley & Bartlette, 2002), which is instead determined by the timing and salience of individual surface events: Lerdahl and Jackendoff's *phenomenal accents*. In this paper, I am concerned with that aspect of phenomenal accent known as durational accent.

2. SOME DEFINITIONS

Durational accent depends on the IOIs that follow and precede each onset; for the sake of interdisciplinary readability, I avoid the possible alternative term "IOI accent". Durational accent does not normally depend on the physical or perceived *duration* of a tone, which can be shorter (in staccato) or longer (in overlapping legato) than its IOI; the corresponding accents may be referred

ISBN 3-931852-67-9
ISSN 1617-6847

to as *articulatory*. In performance, however, duration and legato play important roles in the communication and disambiguation of metrical and grouping structures. For example, the third movement of Brahms's third symphony begins with a 'cello melody whose longest tones are much longer than the IOIs between the phrases. The notated structural interpretation is conveyed to the listener by means of expressive crescendos during the long tones that are marked explicitly in the score and emphasize their legato connection to following tones; and by cutting short the tones at the ends of phrases.

This example shows how important it is to distinguish between *immanent* accents (evident from the score) and *performed* (expressive, *agogic*) accents; expressive performance often involves the reinforcement of the immanent accents of a given structural interpretation by means of performance accents (Friberg & Battel, 2002; Parncutt, 2003). The present paper is concerned primarily with *immanent durational* accents. These correspond to the durational accents of western music theory (e.g. Lester, 1986; Roeder, 2001; see also Lee, 1991; Steedman, 1977), and depend only on the note values (e.g., half- versus quarter-notes) and rest values in musical scores. These also correspond to *psychological categories* that influence both perception and performance (Clarke, 1987; Desain & Honing, 1989): temporal ratios can vary considerably without affecting the corresponding perceptual (and notational) categories.

In everyday musical discourse, the word *accent* tends to mean *dynamic* accent: attention is drawn toward a musical event simply by playing it louder (cf. Dahl, 2000; Windsor, 1994). Like durational accents, dynamic accents can be either immanent (notated) or performed. But in everyday rhythms, durational accent is more important than dynamic accent. In other words, the temporal structure of a simple rhythm tends to be more perceptually salient than its dynamic structure:

- An increase of 2 dB in the level of a tone in an isochronous sequence is sufficient to produce a dynamic accent (Thomassen, 1982), but an increase of about 4 dB is needed to produce a dynamic accent that is strong enough to balance a durational accent (Povel & Okkerman, 1981).
- Young children reproduce temporal structures of musical rhythms more reliably than dynamic structures (Gérard & Drake, 1990).
- Timing has a greater effect on auditory streaming than loudness (cf. Bregman, 1990; Roeder, 2001).
- A number of pulse- and meter-finding models have been developed that consider IOI but not loudness – but not the reverse.
- Durational accents have a greater effect on expressive timing and loudness than do melodic and metrical accents – at least for relatively simple musical patterns such as those performed by the pianists of Drake and Palmer (1993).

The main point of the present contribution is to separate immanent durational accents into two kinds, *predurational* and *postdurational,* and to clarify their separate contributions to overall durational accent. "Predurational accents" (in music theory: "durational accents") precede long IOIs, mark downbeats, and explain why notes at the start of musical measures tend to have greater note values than the local average (Longuet-Higgins & Lee, 1982; Parncutt, 1994; Snyder & Krumhansl, 2001; Thomassen, 1982; Vos, 1977). According to this logic, the strongest predurational accents fall on the final events of pieces of music.

The accents that mark the beginnings of phases are called *structural accents* (Lerdahl & Jackendoff, 1983) or *grouping accents* (Drake & Palmer, 1993). In the present approach, these accents are also called *predurational*. Similarly, accents at the ends of phrases or groups are also *postdurational*.

3. AGOGIC ACCENT

Riemann (1884) defined agogic accent in relation to *tempo rubato* - the local accelerations and decelerations that accompany crescendos and decrescendos. For example, at the start of a phrase a performer might simultaneously increase both loudness and tempo and then dwell on the most important tones of a phrase. There is a general tendency for musicians tend to "steal" a little extra time in the vicinity of important events (Crelle, 1823). In modern North American music theory, the term "agogic accent" usually refers only to the lengthening of important tones in performance. But it is clear from musical experience and psychological research (e.g., Shaffer, 1981; Sloboda, 1983; Todd, 1985; Vos, 1977) that performers direct attention to important events by lengthening *and delaying* them, i.e. by lengthening the following *and* preceding IOIs – consistent with Riemann's approach.

Why does this kind of rubato produce an accent? Agogic accents occur in metrical contexts that generate listener expectations that future downbeats will occur at certain times. Delaying these time points attracts attention to them by heightening expectation, and lengthening the corresponding sound events gives them more weight. In both cases, the listener has more time to perceive the event. The extra time may be psychologically associated with louder sounds that typically require larger production movements than softer sounds (Dahl, 2000).

4. PREDURATIONAL ACCENT

In Parncutt (1994), I investigated the dependency of pulse (beat) perception on predurational accent in simple metronomic rhythms performed at different tempi. The effect tended to be greater at faster tempi. For example, in a simple march rhythm (a cycle of 4 beats divided into 2+1+1), the tendency for listeners to hear the downbeat preceding the longer IOI increased as tempo increased (see also Oberfeld, 2000). Thus, predurational accent is non-linearly related to IOI: it increases with IOI for short IOIs and saturates for long IOIs. This relationship may be modeled by an exponential function of the form $f(\mathrm{IOI}) = 1\text{-}e^{-\mathrm{IOI}/T}$ with a time constant T of about one second.

ISBN 3-931852-67-9
ISSN 1617-6847

In a cognitive-psychological approach, predurational accent can be related to *echoic memory* (Massaro, 1970); it allows the sounds of individual speech phonemes to be stored in a buffer before being categorized and passed to short-term or working memory for further processing. Echoic memory is not memory in the everyday sense of remembering a phone number; like computer memory, it does not involve conscious attention or cognitive processing. Echoic memory has been assigned various names and durations, mostly in the range 0.5 to 2 seconds (for sources see Parncutt, 1994). According to the *proceduralist* assumption (Crowder, 1993), the duration of memory (regardless what kind) is not fixed but depends on the mental activity, its function, and the kind of thing being memorized. It follows that the duration of any kind of memory depends on experimental method.

Why does the phenomenal accent of a rhythmic event depend on the following IOI? A cognitive explanation is that the perceptual salience of a tone is reduced by presenting a second tone within the time-span of echoic memory, interrupting the processing of the first (Massaro, 1970; Povel and Okkerman, 1981). Predurational accent may also depend on the physiology of the auditory system (Todd, 1994; Todd, O'Boyle, & Lee, 1999) and on familiarity with typical auditory environments, in which loud sounds mask softer sounds that immediately precede and (especially) follow them (Parncutt, 1994).

5. POSTDURATIONAL ACCENT

Rhythmic patterns typically last from 2 to 5 seconds (Fraisse, 1982). The temporal gaps separating these patterns determine their segmentation into phrases (Handel, 1981). Postdurational accents mark the start of new phrases, drawing the listener's attention to the continuation of music that might otherwise have stopped – just as a stream of speech may or may not stop at the end of a sentence. Just as the salience of a *pre*durational accent appears to depend on the period of time during which an event is processed undisturbed, *post*durational accent may depend on the period of time during which the *previous* event was processed, and during which the previous phrase is abstracted into higher-level structures (Martin, 1972). The longer this processing lasts, the greater the effect of surprise or the shift of attention when it is interrupted. According to this idea, of the two kinds of durational accent, only postdurational accent depends on attention. It is therefore logical to model it in terms of the *psychological present* (Fraisse, 1982) or *working memory* (Baddeley, 1992). Michon (1978) defined the psychological present as "a time interval in which sensory information, internal processing, and concurrent behavior appear to be integrated within the same span of attention" (p. 89). "It has a width that is highly variable and that seems to have an upper limit of 7 or 8 sec, although its average seems to be of the order of 2 or 3 sec." (p. 92).

Existing approaches to structural or grouping accents (Lerdahl & Jackendoff, 1983; Drake & Palmer, 1993) are *top-down*: the accents are regarded as products of the cognitive process of grouping, their salience depending on the depth of embedding of the group boundaries in a hierarchical structure (cf. Todd, 1985). Here, I assume a *bottom-up origin* for structural accents: they are initially postdurational and determined by the IOIs that precede them, which simultaneously determine segmentation and grouping. The groups then become embedded in cognitive hierarchies whose structure leads to a reassessment of the strength of the grouping accents. A model of the salience of grouping accents should take into account both bottom-up and top-down aspects.

6. MODELING DURATIONAL ACCENT

In a quantitative model of rhythm perception, pre- and postdurational accents may be combined to produce an overall estimate of durational accent. This may then be combined with other phenomenal forms of accentuation such as changes of loudness and timbre. The resultant temporal pattern of accents may then be input to a pulse-finding algorithm that operates within the confines of the psychological present but also accounts for hysteresis.

The relationship between pre- and postdurational accents has been investigated by e.g. Drake and Palmer (1993), Penel and Drake (2001), Povel and Essens (1985), and Oberfeld (2000). Predurational accent is important for shorter IOIs, and postdurational for long IOIs. This relationship may be parsimoniously modeled by exponential functions with different time constants and asymptotic maxima. These parameters can be set so that predurational exceeds postduration accent for shorter IOIs and the relationship reverses for longer IOIs.

Such a model could be tested by comparing predictions with the results of either performance or perception experiments. In the case of performance, a drummer might perform a rhythm while the experimenter measures the loudness or force of the taps. A perceptual experiment might involve EEG measurements as a means of estimating the perceptual salience of rhythmic events in real time.

7. REFERENCES

1. Baddeley, A. (1992). Working memory. *Science, 255*, 556-559.
2. Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.
3. Brown, J. C. (1993). Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America, 94*, 1953-1957.
4. Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 19-33). Stockholm: Royal Swedish Academy of Music.
5. Crelle, A. L. (1823). *Ausdruck und Vortrag für Fortepiano-Spieler*. Berlin: Maurer'sche Buchhandlung.
6. Crowder, R. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), *Thinking in sound* (pp. 113-145). Oxford: Clarendon.
7. Dahl, S. (2000). The playing of an accent – Preliminary observations from temporal and kinematic analysis of percussionists. *J of New Music Research, 29*, 225-233.

ISBN 3-931852-67-9
ISSN 1617-6847

8. Desain, P., & Honing, H. (1989). The quantization of musical time: A connectionist approach. *Computer Music Journal, 13* (3), 56-66.

9. Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10*, 343-378.

10. Fraisse P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *Psychology of music* (pp. 149-180). NY: Academic.

11. Friberg, A., & Battel, G. U. (2002). Structural communication. In R. Parncutt & G. E. McPherson (Eds.), *Science and Psychology of Music Performance* (pp. 199-218). New York: Oxford University Press.

12. Gérard, C., & Drake, C. (1990). The inability of young children to reproduce intensity differences in musical rhythms. *Perception and Psychophysics*, *48,* 91-101.

13. Goto, M. (2001). An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research, 30*, 159-171.

14. Handel, S. (1973). Temporal segmentation of repeating auditory patterns. *J of Experimental Psychol, 101,* 46-54.

15. Handel, S. (1981). Segmentation of sequential patterns. *Journal of Experimental Psychology: Human Perception and Performance,* **7**, 41-55.

16. Krebs, H. (1987). Some extensions of the concepts of metrical consonance and dissonance. *Journal of Music Theory, 31,* 99-120.

17. Large, E. W. (2001). Periodicity, pattern formation, and metric structure. *J of New Music Research, 30,* 173-185.

18. Lee, C.S. (1991). The perception of metrical structure. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 59-127). London: Academic.

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

20. Lester, J. (1986). *The rhythms of tonal music*. Carbondale and Edwardsville: Southern Illinois University Press.

21. Longuet-Higgins, H. C., & Lee, C. S. (1982). Perception of musical rhythms. *Perception, 11,* 115-128.

22. Martin, J. G: (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79*, 487-509.

23. Massaro, D.W. (1970). Preperceptual auditory images. *Journal of Experimental Psychology, 85,* 411-417.

24. Oberfeld, D. (2000). Perception of accents in sequences of long and short inter-onset-intervals. *Psychological Research*, Special issue: Rhythm Perception & Production.

25. Parncutt, R. (1987). The perception of pulse in musical rhythm. In A. Gabrielsson (ed.), *Action and Perception in Rhythm and Music* (pp. 127-138). Stockholm: Royal Swedish Academy of Music.

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

27. Parncutt, R. (2003). Accents and expression in piano performance. In W. Auhagen, B. Gätjen, & K. W. Niemöller (Eds.), *Systemische Musikwissenschaft: Festschrift Jobst P. Fricke.* Frankfurt: Peter Lang.

28. Penel, A., & Drake, C. (2001). Rhythm in music performance and perceived structure. In P. Desain & L. Windsor (Eds.), *Rhythm Perception and Prodn* (pp. 225-232). Lisse: Swets.

29. Povel, D.J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2,* 411-440.

30 Povel, D.J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception and Psychophysics, 30,* 565-572.

31. Riemann, H. (1884). *Musikalische Dynamik und Agogik.* Hamburg: D. Rahter.

32. Roeder, J. (2001). Pulse streams and problems of grouping and metrical dissonance in Bartók's 'With Drums and Pipes'. *Music Theory Online*, 7.1.

33. Shaffer, L.H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13,* 326-376.

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

35. Snyder, J., & Krumhansl, C. L. (2001). Tapping to Ragtime: Cues to pulse finding. *Music Perception, 18*, 455-489.

36. Steedman, M.J. (1977). The perception of musical rhythm and metre. *Perception, 6,* 555-569.

37. Temperley, D., & Bartlette, C. (2002). Parallelism as a factor in metrical analysis. *Music Perception, 20,* 117-149.

38. Thomassen, M. T. (1982). Melodic accent: experiments and a tentative model. *Journal of the Acoustical Society of America, 71,* 1596-1605.

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

40. Todd, N. P. M. (1994). The auditory "primal sketch": A multiscale model of rhythmic grouping. *Journal of New Music Research, 23*, 25-70.

ISBN 3-931852-67-9
ISSN 1617-6847

41. Todd, N. P. M., O'Boyle, D. J., & Lee, C. S. (1999). A sensory-motor theory of rhythm, time perception, and beat induction. *Journal of New Music Research, 28*, 5-28.

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

43. Windsor, W. L. (1994). Dynamic accents and the categorical perception of metre. *Psychology of Music, 21*, 127-140.

ISBN 3-931852-67-9
ISSN 1617-6847