Temporal expectations and their violations

Ladinig, O.

Publication date
2009

Chapter 1

Introduction

1.1 The role of time in music listening

The term that is most commonly used to refer to the temporal dimension of music is *rhythm*. As Fraisse (1982) pointed out more than two decades ago, "the task of those who study rhythm is a difficult one, because a precise, generally accepted definition of rhythm does not exist. The difficulty derives from the fact that rhythm refers to a complex reality in which several variables are fused" (p. 149). His observation seems still to be valid, considering that even today authors still quote this remark, and establish their own definitions of rhythm that focus on the component that is relevant to their particular research question. For example, Patel (2008), after stating that "Unfortunately, there is no universally accepted definition of rhythm" (p. 96), defines rhythm as "the systematic patterning of sound in terms of timing, accent, and grouping" (p. 96). On the most basic and purely physical level, musical rhythm can be described in terms of inter-onset-intervals (IOIs), which are distances of onsets of sounding events along the temporal dimension (Clarke, 1999; Fraisse, 1982). A second parameter of rhythm, though usually seen as less important, is the duration of events, defined as the temporal distance between event onset and offset.

Not surprisingly, physical features are not alone responsible for what is perceived by a listener. The cognitive system tries to organise input in order to facilitate processing and memory. Thus the same sequence of events can be perceived differently - by different listeners, and even by one and the same listener at different times and in different contexts (Desain & Honing, 2003; Repp, 2007). For example, the context in which a rhythm is presented can result in the subjective emphasis of different events. In certain instances a sequence of events will receive subjective accentuation of every second or fourth event; in other instances every third or sixth event will receive subjective emphasis (see discussion of meter below). Furthermore, tempo is a crucial factor and determines if lower- or higher-order relationships between events are established. Notwithstanding these factors that can cause different interpretations of the same rhythmic pattern, the cognitive system is able to abstract invariances from a variety of features and can categorise two rhythms as being the same even if they are played in different tempi, or if they have different expressive timing.

One crucial mechanism that allows us to hear a sequence of events as a rhythm is the tendency to perceive a beat in a pattern of events, usually referred to as *beat induction*. (For a recent overview of research see Patel, 2008.) The ability to perceive a beat seems to be universal in humans and enables them to entrain to music, and to coordinate perception and behaviour. Although some animals show periodic behaviour and are able to produce relatively isochronous (i.e., equally spaced) sounds, it was for a long time believed that they are not able to synchronise with an audio signal, or to pick up regularities from a signal as complex as a musical rhythm. (For the recently fast growing discussion on that subject see Large, Velasco, & Gray, 2008; Patel, Iversen, Bregman, & Schulz, 2009; Schachner, Brady, Pepperberg, & Hauser, 2009.) Interestingly, not all occurrences of a perceived beat have to coincide with an actual event onset, but some can also be perceived in a moment of silence, a so-called *loud rest* in a rhythm. If people are asked to produce beats at a tempo they like (called spontaneous tapping rate, personal tempo, or referent level), the distribution of inter-onset-intervals produced by adults is usually centred around 600 ms (Drake, Jones, & Baruch, 2000; Fraisse, 1982). In this region listeners are also most sensitive in making perceptual assessments - for example, in judging slight tempo differences between two rhythmic patterns (Drake et al., 2000).

Perceiving a beat in a rhythm means to perceive certain events as accented or *salient* events, which do not necessarily have to be (but of course can be) physically accented. Besides an optional physical accentuation, the perception of beats is based on subjective accents that can arise solely from the listener's interpretation of a particular rhythmic structure. Lerdahl and Jackendoff (1983) distinguish three types of accent: accents that arise from the musical surface, such as tone duration or intensity (phenomenal accents), accents that arise from the musical structure, such as boundaries introduced by musical phrases or harmonic progressions (structural accents), and accents that themselves arise from the perception of a beat (metric accents). The first two kinds of accent are physical accents, since they are based on physical cues in the music. The last kind of accent, metric accents, can be seen as subjective accents, since often no physical cues are apparent in metrically accented events.

It is generally believed that based on the perception of subjective accents, each event of a rhythm is perceived as belonging to one particular level of salience, and beats are perceived as having higher salience than the events that occur between them, which in this dissertation are termed *subbeats*. This cognitive tendency of creating a structure with different levels of salience can be observed even with very simple rhythms, such as the ticking of a clock or metronome, where every other event (more rarely, every third or fourth) often receives a metric accent

(Bolton, 1894; Brochard, Abecasis, Potter, Ragot, & Drake, 2003). This phenomenon is termed subjective metricisation (London, 2004). In most common Western rhythms, subbeats (the weaker elements) divide the beat periods into two, three, or four equal intervals.

When at least two periodic levels of perceptual salience can be distinguished, as in the case of subjective metricisation, one speaks of meter perception or *meter induction* (Yeston, 1976). Lerdahl and Jackendoff (1983) state that "fundamental to the idea of meter is the notion of periodic alternations of strong and weak beats" (p. 19). (It should be noted that this terminology differs from the one used in this dissertation. In the present context, 'strong beats' are simply termed 'beats', and 'weak beats' are termed 'subbeats'.) The respective structural levels to which events are perceived as belonging are characterised by their *metric salience*. (Metric salience has to be distinguished from *event salience*, which depends on all three types of accent.) Musicological models of meter (Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1984) assign metric salience to *metric positions* in a hierarchical and recursive way, by dividing the musical measure into equal subparts: The more subdivisions have to be made to assign a position to a structural level, the lower the salience of an event occurring in this position. The first position, usually termed the downbeat, is assigned the highest possible salience of a measure, and the position right after it is by definition always one of the positions with the lowest possible salience. The musical measure itself is, of course, a notational representation of assumed or intended metric units, usually comprising several beats.

1.2 Formation and violation of temporal expectations

Beat induction exemplifies the active role of the listener in the interpretation of rhythm, which involves assigning different perceptual saliences to certain events, even if they are physically identical. This active role becomes even more striking if we consider the phenomenon of a loud rest, which is caused by an expectation for a particular position to have an event, which then does not happen. Many other examples can be found that show that listening to music is not purely passive. What makes perception active is that the cognitive system constantly makes predictions about future events, which amounts to having expectations. One crucial element in our appreciation of music is the balance between expectation fulfilment and violation (Berlyne, 1970; Huron, 2006; Meyer, 1965; Narmour, 1990). Highly predictable music might not be interesting for long, since it is too simple. Music that violates too many expectations will be too complicated for the average listener and sound random, thus also not being perceived as interesting.

Foundations for the study of rhythmic expectations in particular were laid by Mari Riess Jones and colleagues (Jones, 1976; Jones & Boltz, 1989; Large &

Jones, 1999), and are being further developed with neuroscientific methods. (For an overview of research see Zanto, Snyder, & Large, 2006.) Expectation violations result from the relation of event patterns to listeners' underlying cognitive schemata. The more expectations are violated, the higher the perceived complexity of the music. Events with a higher salience are expected more strongly, are better memorised and recalled, and receive the most immediate attention. Consequently the absence of salient events or the disproportionally high occurrence of non-salient events will lead to the impression of complexity (Fitch & Rosenfeld, 2007). In other words, perceived complexity is a good indicator of underlying expectations determined by the metric salience of events.

No generally accepted theory of rhythmic complexity exists yet. (For an overview see Streich, 2007.) However, Smith and Honing (2006) showed that the concept of *syncopation* developed by Longuet-Higgins and Lee (1984) can explain a considerable part of perceived complexity as reflected in listeners' judgments. Admittedly, the concept of complexity can not be entirely reduced to syncopation. Other factors like note density, tempo variation, absence of periodicity and repetition certainly play a role in perceived complexity. However, if as many of these factors as possible are held constant, strength of syncopation may explain the perception of rhythmic complexity. Syncopation is used throughout this dissertation as a way of operationalising complexity of rhythmic patterns. For musicians, syncopation has a relatively clear meaning: a subjective accent in a metrically unaccented position, or a rest in a metrically accented position. A formal definition is given by Longuet-Higgins and Lee (1984), who specify also which parts of a musical measure are accented, to what degree, and how strong a certain syncopation will be. The model of syncopation by Longuet-Higgins and Lee is based on the specification of metric salience for each event in a rhythm by hierarchically and recursively subdividing the measure into equally sized subparts. The metric saliences are then in turn used to calculate the strength of syncopation: A syncopation occurs if a sounding event occurs in a position that has lower metric salience than an immediately preceding position in which no event occurs. The strength of a syncopation is calculated as the difference in metric salience of those events.

The concepts discussed so far (beat, meter, syncopation) presume hierarchical processing of events. Another factor relevant to rhythm perception that was considered in the course of this dissertation is *serial position* within a relatively short rhythmic sequence (Martin, 1972; Jongsma, Desain, & Honing, 2004). The importance of events in comparison with each other may not only be based on hierarchical metric division of the musical measure, but also be related to their absolute placement within one sequence. Effects of serial position are known mainly from memory research (Acheson & MacDonald, 2009; Ebbinghaus, 1885), with events at the beginning and at the end of a sequence being recalled easier (termed *primacy* and *recency* effects, respectively). In this thesis serial position was considered as an additional factor that may determine salience of rhythmic events in a musical measure. The hypothesis was that events at the beginning and at the end of a musical measure would have a higher salience than events in the middle. These issues are addressed in Chapter 4.

1.3 Competence in music listening

There are ways to assess and classify someone's ability regarding music performance. One obvious source of information is the amount of time spent on practice and study of the particular skill in question, as assessed by questionnaires. Less established, however, is a way to assess *listening competence*, or the various capacities for processing music. In their recent review of musical capacities that do not depend on formal music training, Bigand and Poulin-Charronnat (2006) define listening competence as "perceiving the relationships between a theme and its variations, perceiving musical tensions and relaxations, generating musical expectancies, integrating local structures in large-scale structures, learning new compositional systems and responding to music in an emotional (affective) way" (p. 100) as examples of processing musical structure that underlies the musical surface. More systematic research has to be done to evaluate if listening competence grows mainly as a result of formal music training, or if extensive exposure to music can be sufficient to acquire high levels of musical competence via implicit learning.

Most measures used to assess a person's level of expertise are based on the years and intensity of continuous formal music education and practice, as well as on the initial age when the training started. An evaluation of measures used in previous research can be found in Ollen (2006), who reports that 'musical sophistication' is mainly acquired through formal music training and can be assessed by questions about the duration and intensity of the training, the age when the training started, the involvement in music theory courses, and the number of concerts attended on average. However, the question remains if these measures assess someone's competence in music listening. Apart from formal music training, every listener accumulates implicit knowledge based on extracting statistical regularities from simple exposure to music. Ethnomusicological and cross-cultural investigations consider the fact that the culture and musical tradition a listener is exposed to influence his or her perception and cognition, and effects of exposure to different musical traditions are commonly acknowledged. However, effects of exposure can also exist on a smaller scale, within one and the same culture, provided by the many different musical styles or genres available. Despite the fact that musical styles and genres from within one music tradition share basic tonal and rhythmic conventions, variations can be found on a more subtle level, for example in the use of timbre (e.g., acoustic vs. electronic music in contemporary Western culture) or the typical conventions of expressivity specific to different musical styles (for example drum rolls around or after the second snare sound in Breakcore, or the *pump up*, a rise of the key by one step, after a chorus in Pop music). Regarding rhythm, *expressive timing* refers to subtle variations in performance along the temporal dimension, for example 'tempo rubato', the local slowing down and speeding up (Repp, 1990, 1995), or playing slightly before or after the expected time of a note (Desain & Honing, 2003). Unusual expressive timing can make a particular performance distinctive. In Chapter 3 of this thesis, sensitivity of listeners to expressive timing in familiar and unfamiliar styles is assessed as a form of implicit learning.

Removing all effects of formal music training as well as long-term exposure would theoretically allow us to study innate and universal mechanisms regarding the perception and cognition of music. Though impossible to achieve this in laboratory settings with adult participants, in recent years the opportunity to study newborn infants has arisen. Such research can address questions about the origins of music, and about the contributions of nature and nurture to music perception and cognition. Contrary to what has long been believed, human minds do not seem to be tabulae rasae at the time of birth. Recent studies with babies a few days old have revealed that certain perceptual styles and cognitive predispositions govern our perception from very early on (Trehub, 2003; Winkler et al., 2003). Chapter 6 in this dissertation reports an attempt to determine if a basis for the perception of metric salience can be found already in newborn babies.

1.4 Research methods used in this thesis

This thesis reports research that employs both behavioural and electrophysiological measures. One reason for using a variety of methods is the need to assess processes occurring with and without the benefit of attention. Even though it is fruitful to study the effects of different foci of attention on perception and cognition, it is also highly informative to construct experimental conditions that lead to a withdrawal of attention from the material under study, in order to gain insights into automatic processes. Behavioural methods were used to obtain listeners judgments (Chapters 3 and 4) and to measure processing speed and discrimination accuracy (Chapter 5). In the research of Chapters 5 and 6, electrophysiological measures were used to investigate parts of the same processes in adults and infants, respectively, without the benefit of subjects' attention. The analysed component of the induced event-related potentials was the Mismatch Negativity (MMN), which in most cases is not modulated by different attentional states (Sussman, 2007). The next section gives a brief overview of the relevant research methods.

1.4.1 Behavioural methods

One behavioural method used in this dissertation was to collect listeners' judgments, which give indications about conscious assessments and choices regarding the presented stimuli. Judgments are based on reflective thoughts, and there exists the possibility to revise initial decisions before a final answer is reported. In the research described in Chapter 3, listeners made a decision in a two-alternative forced choice task according to the perceived naturalness of two stimuli, and in the research described in Chapter 4 listeners ranked elements within a set of stimuli according to the subjective degree of complexity.

A complementary approach to collecting judgements is to measure listeners' processing speed. Processing speed can be assessed by measuring the reaction time (RT) in experimental tasks. The speed of detecting a deviant stimulus reflects perception and processing while the participant is in an attentive state (Chapter 5). Reaction time is the duration between the presentation of a stimulus and the respective observable response. A fast reaction time can indicate that a deviation was stronger, or that a stimulus was more expected.

Listeners' sensitivity to differences between events can be expressed with the sensitivity index *d'*. This measure takes the correct responses (hit rates) as well as the incorrect responses (false alarms) into account. The higher the *d'*, the more sensitive is a subject in the detection of deviants or in discriminating different events.

1.4.2 Electrophysiological methods

Complementary to traditional behavioural methods, electrophysiological measurements provide a way to study processing without any behavioural responses. Electroencephalography (EEG) measures electricity from the scalp, which is generated by neuronal activity involved in sensation as well as cognition. EEG is an ideal method when studying temporal processing, since it provides high temporal resolution. However, spatial resolution is rather coarse. Electric activity that can be directly related to an event (a thought or a percept) is called an event-related potential (ERP). ERPs are visible as slow waves starting roughly at 100 msec after the event onset. Based on their time course, such waves can be decomposed into various components that have a specific time of occurrence and direction of amplitude, and are seen as stereotypical indicators of certain processes. In practice, ERPs can only be observed after averaging the time-locked data of many trials, since the level of noise (e.g., stemming from the EEG system itself, from spontaneous brain activity, or from body movements) is far beyond the level of the actual signal that has an amplitude of only a few millivolts.

One important component of the ERP was discovered by Näätänen, Gaillard, and Mäntysalo (1978) and termed the Mismatch Negativity (MMN). The MMN signals the detection of an irregular event in a string of regular events, and occurs

150-250 msec after the onset of the deviation. The common view is that the MMN reflects the process of testing a perceptual model against incoming stimuli, with a violation of the model being reflected in the MMN. The auditory MMN can be observed if the deviating event differs from the regular events in terms of frequency, amplitude, duration, or location. Bigger differences between a standard and a deviating event cause bigger and earlier MMN responses. An MMN response can also be observed in case an expected event is simply omitted. By comparing the latency and the amplitude of MMN responses to two different deviations, one can infer which one was more salient. For an overview of paradigms, results, and implications of research using MMN, see Näätänen, Paavilainen, Rinne, and Alho (2007).

The MMN is not only elicited by simple violations of physical properties. Deviations from higher-order regularities can also be reflected. For example, Saarinen, Paavilainen, Schröger, Tervaniemi, and Näätänen (1992) observed MMN responses to descending tone pairs occurring in a series of ascending tone pairs, while the starting tones of the pairs varied. Paavilainen, Simola, Jaramillo, Näätänen, and Winkler (2001) found responses to violations of rules based on relationships between sound attributes, for example "the higher the frequency of a stimulus, the higher the amplitude". Deviants in this case were either low frequency stimuli with high amplitude, or high frequency stimuli with low amplitude.

In our two studies using the MMN (reported in Chapters 5 and 6) we investigated this ability to detect regularities in abstract patterns, and used a set of rhythms with no omissions or with omissions in low-salience positions as standards and a set of rhythms with omissions in high-salience positions as deviants. If the brain creates hierarchical representations of the rhythmic sequences, omission of the most salient event (the downbeat) is expected to elicit stronger responses from participants than omission of a metrically less salient event.

Another useful feature of the MMN component is that it does not require attention to the task, and in most cases does not even benefit from it. This allows use of this method with sleeping newborns (Chapter 6), or with adults who perform some distraction task (Chapter 5).

1.5 Thesis purpose and outline

The purpose of this thesis was to study temporal expectations indirectly by observing reactions to violations of expectations. Expectations on different structural levels were considered, with expressive timing violations testing small-scale expectations, and event omissions testing larger-scale expectations. Special attention was given to listener-specific variables, namely level of formal music training, exposure, and developmental stage, as well as to the role of attention, using methods that monitor pre-attentive as well as attentive processing.

Chapter 2 is a short commentary on the validity of Internet experiments in music perception research. The use of the Internet for data collection is growing in various areas. Still, many scientific journals are sceptical towards this method, and until recently some had a policy to simply reject submissions that report data gathered on the Internet. We argue that the accessibility of a broad range of participants, the saving of time and money, and the increase of ecological validity together with the reduction of experimenter bias are points in favour of that method of data collection.

Chapter 3 reports an Internet experiment that explored the effects of extensive exposure to a certain musical style and of formal music training on a particular aspect of listening competence, namely the sensitivity to small deviations in performance. Scaling musical performances by speeding them up or slowing them down proportionally scales their expressive timing. Previous research has shown that proportionally scaled expressive timing does not sound natural. Participants, either expert musicians or non-musicians, and either familiar or unfamiliar with certain musical styles, were to distinguish between a tempo-transformed and a non-transformed musical performance of the same piece by focusing on the expressive timing. The hypothesis was that formal music training would not be a strong predictor of performance, but that the crucial factor would be the familiarity with a particular musical genre.

Chapter 4 describes an attempt to look at differences between musicians and non-musicians regarding the hierarchical depth of meter perception, expressed in terms of metric saliences. Also based on an Internet experiment, listeners rated the perceived complexity of various rhythms containing omissions in more or less salient positions. The degree of perceived complexity was expected to indicate the degree of expectation violation based on the underlying metric salience. In addition to metric processing, the effect of serial position was studied.

An alternative approach to study metric salience is reported in Chapter 5. Adult non-musicians were required to detect rhythms with two different strengths of syncopation within a string of non-syncopated rhythms, and their reaction time and discrimination sensitivity were measured. Electrophysiological data were gathered for the same stimuli, for two different attentional conditions. The hypothesis was that listeners would be faster and more accurate in behaviourally detecting stronger compared to weaker syncopations, and that stronger MMN responses would be observed for stronger compared to weaker syncopations.

The study reported in Chapter 6 employed a reduced version of the abovementioned paradigm to study the highest level of meter violation, i.e., downbeat omission, in sleeping newborn infants, to see whether expectations based on metric salience are already active at a very early stage in human development.

The dissertation concludes with a discussion and an outlook on future research.

All chapters contain material that is either already published or has been peer-reviewed and is now under revision. The references to published or submitted articles are given below the respective chapter titles. In cases where articles require additions or corrections, these are reported as *Notes* at the end of the text of each chapter.