The perception of accents in pop music melodies

Martin Pfleiderer  
Institute for Musicology,  
University of Hamburg,  
Hamburg, Germany  
martin.pfleiderer@uni-hamburg.de

Daniel Müllensiefen  
Department of Computing, Goldsmiths College, University of London, London, UK  
d.mullensiefen@gold.ac.uk

ABSTRACT
The perception of accents in melodies was explored in a listening experiment with pop music melodies. 29 subjects with long-term music experience listened to short melodies presented as audio excerpts from pop tunes or as single line MIDI melodies. Their task was to rate every note of every melody for their perceived accent value on a 3-point-scale. The ratings were checked for consistency. The most consistent ratings were averaged and then compared to the results of 34 accent rules taken mainly from the literature on melodic accent perception. Two statistical procedures (linear regression and regression trees) were subsequently used to determine the optimal combination of rules to predict the ratings. Both models were later tested on a separate data set. Results indicate that for the perception of monophonic melodies several Gestalt rules are employed by the listeners, while the perception of melody accents in pop music excerpts is largely governed by the strong metrical beats that are in most cases outlined very clearly by accompanying instruments as well as by syncopation. Finally, the implications of the experimental and statistical methods for future research on melody and rhythm perception are discussed.

Keywords
Melody perception, accent perception, Gestalt rules, statistical modelling, rhythm perception, beat induction

INTRODUCTION
There are numerous attempts to explain the perception of accents in a melody by a simple rule or a combination of rules that are based on principles of Gestalt perception. Quite a lot of these rules have been empirically tested in listening experiments (e.g. Thomassen, 1982; Boltz & Jones, 1986; Jones, 1987; Jones, Summerell & Marshburn, 1987; Monahan, Kendall & Carterete, 1987; Jones, 1993; Jones & Ralston, 1991; Boltz, 1999). Results of these studies almost always indicate that subjects apply Gestalt-like rules in their perception of accents in melodies or short note sequences. The present study was motivated by several shortcomings that are common to the above cited studies.

Firstly, in most experiments in this field monophonic sequences of notes are used that do not come from existing tunes, but are assembled (“composed”) by the experimenters (e.g. Eiting, 1984). This makes experimental stimuli easy to control, because they can be manipulated so that a subset of the sequence shows a certain trait or falls in a certain category in terms of a Gestalt-law (e.g. large jump in melodic contour) and another subset doesn’t. The influence of that specific trait on melodic accent perception can then be assessed easily by statistical tests. But at the same time the possibility to generalise the experimental results from artificially constructed sequences to real melodies of any style often remains unexplored. In contrast to those artificial stimuli, in our study only “real” pop music melodies are used.

Secondly, to our knowledge no experiment to date has tested the descriptive power of basic Gestalt-rules for melodic accent perception with excerpts from real songs including not only the sung melody but all other ‘natural’ musical components too, like all accompaniment instruments (chord instruments, bass, drums). To explore this issue we tested the subjects with monophonic versions and with audio excerpts of the same melodies.

Thirdly, some studies propose promising models in which several basic perceptual rules are combined operating on different dimensions of melody perception, e.g. melodic contour, intervals, rhythm, meter (Boltz & Jones, 1986; Monahan, 1987; Boltz, 1999). But to our knowledge there hasn’t been a single study that tries to find an optimal combination out of a large set of possible rules. Similarly, there has not been any attempt to find out whether the rules employed in a rule combination model are all of the same importance or weight, or if there are rules that are more important than others in explaining human accent perception.
To create a combination model of rules that are mutually weighted is therefore a goal of the present study.

THE PERCEPTUAL RULES

This study builds on recent work by Müllensiefen & Frieler (2006) and Müllensiefen (2004) which aims at systematizing and completing the set of simple Gestalt-rules of different dimensions for melody perception. In the course of this work 34 rules proposed in the literature (mainly from Monahan et al., 1987; Boltz & Jones, 1986; Boltz, 1999) or derived from our own considerations were implemented in a software toolbox that can be used to analyse monophonic melodies. The output is a set of 34 binary vectors each vector having the same number of elements as the number of notes in the melody. Each binary vector element represents the existence or the absence of a given accent feature at a particular note in the melody. The rules implemented in the toolbox cover many dimensions that are commonly suspected to play a role in the perception of melodic accents. There are rules based on events in the interval dimension (jumps), contour dimension (turning point in melodic direction), rhythm (note durations), grouping (phrase endings and beginnings), harmony (harmonic relation of individual notes underlying chords), and meter (on-beats, syncopation, and overall degree of syncopation). A list of all the rules used for analysis in this study is included in the appendix. A more in-depth discussion on the rule set and its implementation can be found in Müllensiefen and Frieler (2006). The rule set was implemented in a software toolbox named melfeature which was programmed by Klaus Frieler.

As an illustrative example of how this rule based analysis works, figure 1 shows a short melody taken from the song “Cold Cold Heart” by Wet Wet Wet in staff notation and the binary values resulting from the application of three accent rules.

![Melody notation](image)

**Figure 1: Melodic excerpt from song: “Cold Cold Heart” by Wet Wet Wet. Below the staff notation the values of three rules for each note are depicted.**

**longpr**: This rule assigns a value of 1 to all notes with an inter-onset interval (IOI) longer than the IOI of the previous note. For example, the third note is longer than its predecessor, thus it receives a value of 1.

**jumpapt3**: This rule assigns a value of 1 to all notes that are preceded by an interval jump of at least 3 semitones, e.g. the second and third note of the example melody.

**phrasend**: According to this rule every note that is located on a phrase ending is assigned a value of 1. Phrase endings are defined by notes with an inter-onset interval of at least 4 times of the most frequent inter-onset interval (the modus of the IOIs) found in the entire melodic sequence. The last note of a sequence is always counted as constituting a phrase ending. In this example the mode of IOIs is a quaver; so, a phrase ending can be found only on the last note. Just defining a threshold value for the IOI of a note to make a phrase ending is probably the simplest way of segmenting a longer melodic passage into smaller segments. For the application of the phrasend rule more sophisticated models of melodic segmentation could be employed as well, like the ones proposed by Temperley (2001) or Combouropoulos (1998). But informal tests with these models showed unsatisfactory results when applied to pop music melodies. The simple segmentation rule used here could therefore be seen as the least controversial of almost all existing segmentation models that music experts can always agree on.

Having applied the 34 rules to the 15 test songs the results are 34 binary vectors for each song. There are three primary goals of this study:

Firstly, to identify the most important rules that come closest to human accent perception.

Secondly, the definition of a model in which the perceptual rules can be combined. This follows the idea favoured by Boltz and Jones (1986), and by Monahan et al. (1987) that accent sources from different musical dimensions are evaluated simultaneously in human melody perception.

Thirdly, there is still an open question, whether the combination of rules is done by simply adding the values of the individual rules. Therefore we tested a linear model and a tree model which combine the values of the individual rules in different ways. The resulting models should specify the importance or the weight of the individual rules within the model.

METHOD

Participants

29 students of an introductory course in music psychology at Hamburg University participated in the experiment (12 female and 17 male students from 20 to 29 years old, mean age 23.2). All of the subjects showed strong preferences for popular music styles and suggests that they were familiar with the musical idiom of the test items (see below).

Procedure

The experiment consisted of two similar sessions that were conducted two weeks apart. In the first session, the participants were told that the experiment was about the perception of melodies and that the degree of accentuation can differ among the melody’s notes. The actual task consisted of listening to a melody and to mark each note within a graphical representation of the melody on a paper sheet indicating the subjectively perceived degree of accentuation. A pre-test showed that a three-point scale was optimal for this
kind of task, because more scale degrees made the subjects’ decision process considerably slower and more difficult. The three scale steps were ‘not accentuated’ (no mark, coded as 1), ‘accentuated’ (circle around note, coded as 2), and ‘strongly accentuated’ (circle and additional cross, coded as 3). The graphic representation of the melodies shows only the approximate lengths (mainly crotchets and quavers) and the relative pitch heights of the individual notes. There were no rests, bar lines, or staff lines included. If the listening example included vocals, the lyrics were printed below each note (see figure 2).

Figure 2: Graphical representation of a melody (“Cold Cold Heart” by Wet Wet Wet, see fig.1) on the test sheet.

The graphical representations were intended to be easy to follow along while listening to the melody excerpt. At the same time they contain all the necessary elements to fulfill the experimental task and are still quite close to the original transcriptions from which they were derived (see Figure 1). It can not be ruled out that the graphical representation had an effect on the participants’ markings. This should be compared in a future study with a different feedback method.

Each test melody was presented four times. The subjects were advised just to listen to the melody first, to circle the important notes the second time, to make crosses at the very important notes the third time and finally to examine their own ratings the last time the melody was played.

After the instructions a training example was played to the subjects to ensure that the graphical melody representation and the marking scheme were well understood. Subjects could indicate whether they had any problems understanding the concept of melodic accents or the task itself.

Both test sessions consisted of 17 melodies, of which 2 were used exclusively to check for within subject reliability (see below). Silent intervals of ca. 4 seconds were inserted between the four repetitions of each melody while different melodies were separated by 10 seconds of silence. In a pre-test participants reported that for them it was much easier to rate the monophonic melodies than the original excerpts. To lighten task concentration in both test sessions first the original melodies and thereafter the monophonic examples were presented. Each test session had a duration of 25 minutes.

After the first test session subjects had to answer a questionnaire on personal data, musical training, musical preferences and feedback concerning the experimental task. The second experimental session involved the same subjects, but different versions of the 17 melodies.

Stimuli: The melodies

There were two versions of each melody: An excerpt from the audio recording of a popular music song (“original melody”), and a one-voice melody extract (“MIDI melody”) played with the “grand piano”-sound (MIDI patch 1) of the general MIDI device of a PC (Microsoft Software Wavetable Synthesizer with Roland Sound Canvas digital samples). The MIDI melodies were based on the vocal melodies of the recording excerpts transcribed by the authors. In the first experimental session eight original melodies and nine MIDI melodies were presented. In the second session, the subjects listened to the original melodies corresponding to the MIDI melodies they had had heard two weeks before and vice versa, except for reliability items. To test the reliability of the subjects’ ratings, one original melody and one MIDI melody were played in both parts of the experiment. Most of the excerpts were chosen from successful, but not very well known pop hits of the immediate past. In addition, there were some excerpts of recordings of contemporary R’n’B singers, one example of Jamaican ska, one reggae song and one example of Trinidadian calypso (a list of the artists and recordings can be obtained by the authors). Every excerpt was between 10 and 20 seconds long and had a 2- or 4-beat meter with binary phrasing (no “triplet quavers”). Tempo varied between 75 and 155 bpm.

RESULTS

Since we expected the perception of melodic accents to be different for the original melodies and the MIDI melodies, all calculations concerning reliability and model construction were carried out separately for the two data sets.

Within subject reliability

One of the implicit assumptions of modelling human melody perception with algorithmic rules is that for the same stimuli subjects should assign the same accent values to the melody notes on every occasion. For this reason we first assessed the reliability of the subjects’ ratings for the two test items that were played in the same version in both test sessions. We chose an accuracy threshold of at least 85% to include a subject’s rating data in the subsequent model construction. Subjects had to achieve the 85% accuracy on two different measures.

First, their accent ratings of the same test items in the two test session had to result in a Pearson-Bravais correlation coefficient of r>0.5, which is roughly equivalent to a maximal difference between two accent ratings (difference value of 2) on three notes or four notes (15%) in a 20- or 29-note melody respectively.

For the second measure the accent ratings were projected onto a binary scale of accent vs. non-accent ratings by simply converting all values of 3 to values of 2. The ratings of one subject for the same test item in the two experimental sessions had to be identical at least in 85% of the notes.
(at least 17 and 25 notes in the 20 and 29-note melody respectively).

Only the data from the four reliable subjects for the MIDI melodies and seven reliable subjects for the audio items that fulfilled both criteria were included in the construction of the statistical models. The data from the 25 and 22 remaining subjects was saved for model evaluation.

**Between subject reliability**

A second step in data selection was to exclude the rating data of all test items where the reliable subjects disagree in their judgments. In order to do that in an objective way, we employed two accuracy measures to measure the degree of correlation between various vectors of the same length. The first measure is Cronbach’s alpha which is widely used in individual differences as an indicator of how several variables estimate a common but possibly latent magnitude. The second measure is the Kaiser-Mayer-Olkin-measure (KMO; Kaiser, 1974) commonly used to evaluate the global coherence of different variables within a correlation matrix. As such it is often employed to decide the aptitude of a correlation matrix for a subsequent principal components analysis. We set a value of 0.6 as threshold for both measures for the data of the melody to enter the statistical modelling. For the KMO a value of 0.6 is considered to be sufficient (Brosius, 1998).

Of the 15 MIDI melodies (not counting the two melodies that served as reliability items) the reliable subjects agreed on the perceived accents in 6 melodies according to the two measures. Of the 15 original audio melodies 11 have been selected according to our criteria.

Only the ratings of the reliable subjects on the reliable melodies were selected as training data for the two statistical models. The arithmetic mean and the median of the subjects’ ratings were taken to condense the data to an average value for each note.

**Statistical modelling of subjects’ ratings**

We employed two different statistical techniques to model the subjects’ ratings (dependent variable) using binary vectors from the accent rules as independent variables: Linear regression and a regression tree.

**Linear Regression**

Linear regression models the dependent variable as a weighted sum of a set of independent variables. A linear regression equation has the basic form:

\[
\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n,
\]

where \(\hat{y}\) is the dependent value (arithmetic means of training data) to be predicted, \(a_0\) - \(a_n\) are weights to be estimated from the training data and \(x_1\) - \(x_n\) are the independent variables. The mathematical form of this model comes closest to the assumptions made by Boltz and Jones (1986) and others that the accents predicted by different rules combine additively to an overall sum for each note of the melody.

The first step was to apply a selection technique that would leave only those of the 34 independent variables that contribute to the explanatory power of the model as defined by a standard F-test criterion. We used the backwards stepwise elimination procedure that is built into the SPSS regression package. After eliminating those variables that failed to reach the F-test significance there were seven variables left in the model for the original audio melodies (corrected \(r^2 = 0.55\)) and another seven variables for the MIDI melodies (corrected \(r^2 = 0.54\)). As a second step a coefficient estimation was performed for the variables in the models. Due to space limitations we can only show here the selected variables, their standardized beta-weights and corresponding p-values in table of the linear model for the original audio melodies in table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta coefficient (unstandardised)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>1.110</td>
<td>.000</td>
</tr>
<tr>
<td>BEAT1</td>
<td>.524</td>
<td>.000</td>
</tr>
<tr>
<td>SYNK2</td>
<td>.330</td>
<td>.000</td>
</tr>
<tr>
<td>PHRASEND</td>
<td>.259</td>
<td>.000</td>
</tr>
<tr>
<td>PEXTRMF</td>
<td>.195</td>
<td>.001</td>
</tr>
<tr>
<td>LONGMOD</td>
<td>.190</td>
<td>.000</td>
</tr>
<tr>
<td>JUMPAFT4</td>
<td>.183</td>
<td>.000</td>
</tr>
<tr>
<td>BEAT13</td>
<td>.155</td>
<td>.021</td>
</tr>
</tbody>
</table>

The standardized beta-weights give an indication of the importance of the variables in the model. As can be seen the rule **beat1** has the highest beta-weight and **beat13** is also comparatively strong. According to these rules notes that fall on the first beat of a bar and on the first or third beat of a bar tend to be perceived as accented when people listen to real music. But according to the weight on **synk2** strongly syncopated notes, i.e. notes that are shifted from the first or third beat of the bar to the offbeat just before, are accented as well. Strong weights are also given to notes on phrase endings (phrasend), notes at extremes of the melodic contour (pextrmf), notes that are longer than the most frequent duration in a melody (longmod) and notes that follow a jump of at least 4 semitones (jumpaft4).

A qualitative view at the components of the model for MIDI melodies reveals that most of the rules involved correspond in some way to the rules selected by the tree model explained below.
Regression tree

Although regression and classification trees have been a standard technique in machine learning for more than 20 years they are very rarely used in modelling data from music perception experiments (exception are Müllensiefen, 2004; Kopiez et al., 2006). The idea behind regression trees is to partition a number of cases into a few categories of the dependent variable by a set of independent variables. The process of partitioning is hierarchical and recursive so that the resulting output lends itself very easily to be displayed in a graphical tree structure. Among other convenient features of tree models are the built-in mechanism of variable selection and the ability to deal with missing cases by so-called surrogate variables (see Müllensiefen 2004) for a discussion of the advantages of tree models for research in music perception. We used the CART algorithm as proposed by Breiman et al. (1984) and as implemented in the rpart package of the statistical software environment R (http://www.r-project.org/).

The application of the regression tree procedure to the MIDI melody data (medians of the ratings) lead to the model depicted in figure 4.

![Figure 4: Regression tree model for accent perception in MIDI melodies](image)

If the condition of a rule is fulfilled then the reader should go down the right branch at that particular split. Otherwise the left branch should be followed. The numbers in the final boxes represent the mean of the medians of the accent values for all melody notes that follow each tree path outlined by the rules in the boxes above them. The relative error of this model is 0.326. This means that 67.4% of the variance in the data is explained by the model, which corresponds to the concept of $R^2$ in a linear regression model.

The graphical tree structure is the easiest way to understand the tree model. Nevertheless, an explanation in natural language can also be given.

The rule that partitions the data first is longpr. Shorter notes are then checked to see if they fall on a phrase ending (phrasend). So notes that are shorter than their predecessor and don’t make a phrase ending receive a low accent value of 1.044. Short notes that fall on phrase endings receive an accent value of 2. Going down the right branch at the top of the tree model we follow the path of the long notes. If a long note doesn’t follow an interval jump of at least a minor third (jumpaft3), they receive a relatively low value of 1.346. For longer notes that actually are a second note of such an interval, the overall syncopation level of the melody (i.e. the number of syncopated notes related to the overall number of notes) is checked (syncop; this is the only non-binary rule; the values range between 0 and 1). If the note is not part of a highly syncopated melody, i.e. a third or less of the melody’s notes are syncopated, then it receives a high accent value of 2.5. If this long and part-of-jump note is played within a highly syncopated melody, then its accent value depends on whether it falls on beat 1 or 3 of the bar (beat13). If this is the case, then it gets an accent value of 2 otherwise it receives a value of 1.

This rather lengthy explanation of the graphical model can be summarized: A short note only gets an accent when it falls on a phrase ending. Longer notes get a strong accent if they are the second note of a jump and if the entire melody has few syncopations. If the melody has many syncopations the long notes after large intervals get a moderate accent when they are on beat 1 or 3.

This tree model is very much in accordance with what the joint-accent-hypothesis (Jones & Boltz, 1989) predicts. The right branch of the figure tells us that accents from rhythm and the interval dimensions have to come together, and that – depending on the overall syncopation level – even a metrical accent must be present for the note to be perceived as accentuated.

The tree model we computed for the original melodies was quite similar to the linear model solution in respect to the incorporated rules (i.e. the dominance of beat1), but due to space limitations a detailed explanation is left out here.
**Model evaluation**

To compare the two types of models and to get an estimate of their predictive power on unbiased data, we tested all four models (linear and tree models for original and MIDI melodies respectively) on the data that had been left out so far (i.e. all ratings of the unreliable subjects and ratings for unreliable melodies by the reliable subjects). As an estimator of the predictive accuracy we defined $a$, as the average difference between the value predicted by the model and the median of the subjects’ ratings:

$$ a = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} $$

with $y_i$ being the median of the subjects’ accent ratings for a particular note $i$, $\hat{y}_i$ being the predicted value from a model for the note $i$, and $n = 382$ (i.e. the number of notes of all test songs).

The evaluation results can be seen from table 2 which also contains the accuracy of the trivial model which takes the mean of the training data as a constant prediction value.

<table>
<thead>
<tr>
<th>Model</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIDI linear model</td>
<td>0.303</td>
</tr>
<tr>
<td>MIDI tree model</td>
<td>0.257</td>
</tr>
<tr>
<td>Mean MIDI training data</td>
<td>0.43</td>
</tr>
<tr>
<td>Audio linear model</td>
<td>0.332</td>
</tr>
<tr>
<td>Audio tree model</td>
<td>0.387</td>
</tr>
<tr>
<td>Mean Audio training data</td>
<td>0.504</td>
</tr>
</tbody>
</table>

Table 3 shows that for the original audio melodies the linear model that was outlined above gives clearly the best predictions. On average the difference between the model prediction and the median of the subjects’ ratings is 0.332 accent values (of the original 3-point accent value scale). The tree model is slightly worse and the mean rating from the test data (1.3) still shows a greater difference on average to the subjects’ ratings. Looking at the models for the MIDI data the tree model that was explained in detail above has the best value for $a$ (0.257). But the linear audio model is close behind ($a = 0.303$), while again the mean from the training data as constant value is much worse ($a = 0.43$). So the statistical modelling can be considered as successful as the prediction accuracy of the models is clearly superior to the trivial model of just taking the mean of the training data.

**DISCUSSION**

The results and the experimental and statistical approach we employed in our investigation have several implications for future research on melody and accent perception as well as on rhythm perception and beat induction.

**Methodological advantages of the approach**

Firstly, it must be emphasized that our experimental design is a powerful tool for the empirical exploration of the importance, weight or accentuation of notes in melodies. Most of the subjects participating in the experiment reported that they seldom had difficulties relating the note symbols to the heard tones. They could maintain a good level of concentration during the task (with a median of 7 on a 10-steps rating scale) and they rate the task to be of medium difficulty (median of 5).

Secondly, linear regression and regression trees are both adequate statistical methods leading to convincing and interpretable results. An important advantage of both methods is the possibility of further differentiating the explored accent rules. Several formulations of rules concerning the same musical dimension, e.g. accents after leaps of different magnitude or several accent rules concerning the highest (and lowest) note in the melodic contour, could be tested at the same time). The most adequate of these rule formulations or an appropriate rule combination will be chosen by the statistical procedure according to its predictive power. Going one step further in a possible future experiment, global rules could be formulated that assign a value to all tones of a test melody. By choosing different values for example for the overall degree of syncopation (as we did in this study) or for the complexity of melodies, or by just giving music of different musical styles different values it would be possible to test if these features influence the applied rules, rule combinations or accent values.

Thirdly, as the results of our experiment indicate, there are differences between the perception of isolated, monophonic melodies and the perception of original melodies including their “natural” accompaniment, at least in the case of pop melodies. Whereas the importance of notes in single line melodies were perceived mainly according to Gestalt rules (inter-onset interval, pitch jumps) the most important features determining accent perception of the same melodies in their original context emphasized the meter, which in general is highlighted by the accompaniment of the pop songs, and the syncopations of the vocal melody. Future investigations should explore whether other features of the vocal interpretation of a melody, e.g. timbre, articulation, dynamics, micro-rhythmic phrasing, or semantic meaning of the lyrics influence the accent perception as well; in our experiments we found only weak hints for that hypothesis.

**Single line melodies vs. original song excerpts**

It could be questioned if experiments with single line melodies possess sufficient ecological validity. Because of the overall presence of music from CD- and record players, radio and music television today most melodies, are listened to – and we propose are sung and remembered as well – in a real or imagined accompaniment context. So
even when somebody is humming or singing a melody the singer can be assumed to have the accompaniment on his or her mind.

It has to be strongly emphasized that the regression model for the original melodies in our experiment not only highlights the beat accent rules but syncopated notes as well. If only the notes on beats had been perceived as accented by the listeners, it would suggest the pessimistic assumption that pop listeners today tend to exhibit a rather simplistic perception of pop tunes, in that they might be attracted only by a dull stereotype beat. But the experimental results indicate that this is actually not the case. The weight of a steady metrical frame is balanced by a strong awareness of syncopations common in rock and pop music (see Temperley 1999, Pfleiderer 2006) – syncopations that could only occur within a perceived metrical frame.

Implications for rhythm theory

Finally, our approach leads to more theoretical considerations concerning the role of accents in approaches to meter induction and rhythm perception. As Eric F. Clarke stated several years ago, in empirical work on these issues there is a lack of the exploration “(…) of different kinds of accent and the ways in which accentual and temporal factors mutually influence each other (…). This is unfortunate both because purely temporal models of meter perception (…) are unrealistic in the demands that they make by deriving meter from temporal information alone and because such models tend to project a static and one-dimensional view of meter, rather than the more dynamic and fluid reality that is the consequence of the interplay of different sources of perceptual information (temporal and accentual)” (Clarke 1999, p. 489). In our experimental design both types of accents, temporal accents (IOI and duration) as well as accents resulting from information on pitch, harmony etc., are explored. So we hope to contribute to the construction of more reliable rule systems for meter induction.

Moreover, our approach could lead to an appreciation and empirical evaluation of the rhythm theories of Maury Yeston (1976) and Peter Petersen (1999). According to these theoretical approaches several accent types could constitute different rhythmic strata (Yeston 1976) or rhythmic components (Petersen 1999) that work together – or conflict with each other – to build the more or less complex overall rhythmic structure of a piece of music (see also Pfleiderer 2006). Our experimental design could be used to identify the relevance and relative weight of possible rhythmic strata or components corresponding to the accent types of our accentuation rule pool. Of course, this pool is always open to new accent rule candidates. By taking into account the temporal and non-temporal accents and combining approaches to melody perception and rhythm perception we hope to contribute to an empirical answer to one of the central questions of music research: What is of importance in human experience of music?

ACKNOWLEDGMENTS

Thanks to the 29 participants in the experiment, Steffen Just for entering the data, Marcus Pearce for revising the draft of this paper, and to Klaus Frieler for programming the software toolbox.

REFERENCES


**Appendix: Table of accent rules**

(for detailed description and discussion of the rules see Müllensiefen & Frieler, 2006)

<table>
<thead>
<tr>
<th>Rule Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUMPAFT[3, 4, 5]</td>
<td>Accent on note after a jump of 3, 4 or 5 semitones</td>
</tr>
<tr>
<td>JUMPBEF[3, 4, 5]</td>
<td>Accent on note before a jump of 3, 4 or 5 semitones</td>
</tr>
<tr>
<td>JUMPBEA[3, 4, 5]</td>
<td>Accent on notes before and after a jump of 3, 4 or 5 semitones</td>
</tr>
<tr>
<td>JUMPLOC</td>
<td>Accent on second note of an interval that is at least two semitones larger than its successor and predecessor interval</td>
</tr>
<tr>
<td>SHORTPHR</td>
<td>Accent on second note of melody phrase consisting of only two notes</td>
</tr>
<tr>
<td>PEXTREM</td>
<td>Accent on note where predecessor and successor notes are both lower or higher</td>
</tr>
<tr>
<td>PEXTRST</td>
<td>Same as PEXTREM but filtering for change notes in the definition of Steinbeck</td>
</tr>
<tr>
<td>PEXTRMF</td>
<td>Same as PEXTREM but filtering for change notes in the definition of Müllensiefen &amp; Frieler</td>
</tr>
<tr>
<td>PEXTRSTA</td>
<td>Accent on note following note accented by PEXTRST</td>
</tr>
<tr>
<td>LONGPR</td>
<td>Accent on note starting an IOI longer than predecessor IOI</td>
</tr>
<tr>
<td>LONG2PR</td>
<td>Accent on note starting an IOI at least 2x as long as predecessor IOI</td>
</tr>
<tr>
<td>LONGMOD</td>
<td>Accent on note starting an IOI longer than mode of IOIs in melody</td>
</tr>
<tr>
<td>LONG2MOD</td>
<td>Accent on note starting an IOI at least 2x as long as the mode of IOIs in melody</td>
</tr>
<tr>
<td>SHORTPR</td>
<td>Accent on note starting an IOI shorter than predecessor IOI</td>
</tr>
<tr>
<td>SHORT2PR</td>
<td>Accent on note starting an IOI at most half as long as predecessor IOI</td>
</tr>
<tr>
<td>ENDLOIOI</td>
<td>Accent on note that ends IOI which is at least 2x the mode of IOIs in melody</td>
</tr>
<tr>
<td>BEAT1</td>
<td>Accent on beat 1 of a bar</td>
</tr>
<tr>
<td>BEAT13</td>
<td>Accent on beat 1 and 3 of a bar</td>
</tr>
<tr>
<td>BEATALL</td>
<td>Accent on all beats of a bar</td>
</tr>
<tr>
<td>SYNK1</td>
<td>Accent on note with onset not on any beat of a bar and with IOI extending over the next beat</td>
</tr>
<tr>
<td>SYNK2</td>
<td>Accent on note with onset a quaver before beats 1, 3 of a bar and with IOI extending over next beat 1 or 3</td>
</tr>
<tr>
<td>SYNKOP</td>
<td>Overall syncopation level of melody</td>
</tr>
<tr>
<td>HARMONY</td>
<td>Accent on note that is part of the accompanying harmony</td>
</tr>
<tr>
<td>DISSBEAT</td>
<td>Accent on note on a beat but not part of the accompanying harmony</td>
</tr>
<tr>
<td>TRIAD</td>
<td>Accent on note that is part of implied harmony of the bar</td>
</tr>
<tr>
<td>TRIADPHEN</td>
<td>Accent on note that is part of implied harmony of the bar and ends a phrase</td>
</tr>
<tr>
<td>PHRASEBEG</td>
<td>Accent on phrase beginning</td>
</tr>
<tr>
<td>PHRASEND</td>
<td>Accent on phrase end</td>
</tr>
</tbody>
</table>