Statistical techniques in music psychology: An update

Daniel Müllensiefen

1 Introduction

Music psychology as a discipline has its origins at the end of the 19th century and ever since then, empirical methods have been a core part in this field of research. While its experimental and analytical methods have mainly been related to methodology employed in general psychology, several statistical techniques have emerged over the course of the past century being specific for empirical research in music psychology. These core methods have been described in a few didactic and summarising publications at several stages of the discipline's history (see e.g. Wundt, 1882; Böttcher & Kerner, 1978; Windsor, 2001, or Beran, 2004 for a very technical overview), and these publications have been valuable resources to students and researchers alike. In contrast to these texts with a rather didactical focus, the objective of this chapter is to provide an overview of a range of novel statistical techniques that have been employed in recent years in music psychology research.[1] This overview will give enough insight into each technique as such. The interested reader will then have to turn to the original publications, to obtain a more in-depth knowledge of the details related to maths and the field of application.

Empirical research into auditory perception and the psychology of music might have its beginnings in the opening of the psychological laboratory by Wilhelm Wundt in Leipzig in 1879 where experiments on human perception were conducted, and standards for empirical research and analysis were developed. From the early stages until today, the psychology of music followed largely the topics and trends of

[1] This paper has also been inspired by conversations with Albrecht Schneider in the context of jointly taught seminars on *Advanced Statistical Techniques in Music Psychology* at the Hamburg Institute of Musicology. On these occasions, Albrecht Schneider repeatedly mentioned that it was about time to write an update of the standard textbook *Methoden in der Musikpsychologie* by Böttcher and Kerner (1978). While this paper can hardly be considered a didactical text nor does it describe a firm canon of modern and frequently methods in music psychology, it still might convey an impression of how such an update might look like (if one actually was to write one).

general psychology, each time with a considerable time lag, including psychophysics (psychoacoustics), Gestalt psychology, individual differences, cognitive psychology, and computer modelling. Each field of research is associated with a canon of empirical and statistical methods which are shared among the research community.

In this chapter we will highlight a few selected statistical methods used in more recent empirical studies which might indicate new trends for the analysis of psycho-musicological data. We focus on two fields of research: first, cognitive music psychology, and second, computer models of musical structure with psychological constraints and mechanisms as core parts of the models. The distinction between these two areas is somewhat arbitrary, and is made on the pragmatic grounds that studies in the cognitive psychology of music mainly aim at explaining experimental data from human subjects whereas in computer modelling the music itself is taken as the data that is to be explained. Due to space limitations we will not cover adjacent areas such as the neurosciences of music, music information retrieval, or individual differences (including tests of musical ability). As an engineering science, music information retrieval (e.g. Orio, 2006) has developed a vast arsenal of sophisticated statistical and machine learning techniques, although it has been lamented (e.g. Flexer, 2006) that proper statistical evaluation of algorithms and techniques is often under-represented in this field. Individual differences and personality psychology naturally have a strong connection to test construction and measurement theory, while methods of statistical evaluation in the neuropsychology of music tend to be similar to those in cognitive music psychology, and thus can be considered as being partly covered in this short review. Inspired by neuroscience research is the deployment of artificial neural networks to model cognitive music systems or experimental data. However, the large body of literature describing the use of artificial neural networks in music psychology (see Bharucha, 1987; Desain & Honing, 1992; Tillmann et al., 2003) is beyond the scope of this review.

2 Cognitive psychology of music

2.1 The standard repertoire

Cognitive psychology of music traditionally explains music perception and music related behaviour in terms of mental mechanisms and concepts, such as memory, affects, mental processes, and mental representations. While mental mechanisms and concepts are usually unobservable, it is possible to generate hypotheses about empirically evident phenomena from the assumptions and specifications that the cognitive concepts imply. Therefore, a large amount of empirical work uses inferential statistics to test the validity of cognitive models and their parameters. Standard techniques comprise procedures for univariate comparisons between groups, mainly in their parametric version (uni- and multi-factorial ANOVAs, t-test etc.) and, to a lesser degree, include also robust techniques such as rank-based tests. For relating one dependent to several independent variables on an interval level of measurement, linear regression is a popular statistical method (e.g. Eerola et al., 2002; Cuddy & Lunney, 1995), although rigorous evaluation of the predictive power of the resulting regression models is seldom undertaken. Bivariate correlation, frequency association measures (χ^2), and descriptive statistics are also commonly employed to arrive at an understanding of the relationships between a number of predictors and one dependent variable (a simultaneous analysis of several dependent variables is possible but unusual).

In summary, the main concern of a large number of empirical psychomusicological studies is to identify one or more important (i.e. significant) determinants for cognitive behaviour as measured by a dependent variable. However, the amount of noise in the linear model is usually of less concern. That is, low R^2 values in regression models are not really a matter of concern, and generally little thought is invested in alternative loss functions or measures of model accuracy. Thus, the focus has been mainly on hypothesis testing rather than on statistical modelling. The beneficial side of this 'statistical conservatism' in music psychology is that experimental data and results can be easily compared, exchanged, and replicated. But at the same time, the reluctance to explore the variety of statistical modelling techniques available nowadays for many specific analysis situations, might leave interesting information in some experimental datasets uncovered. A review of these quasi-canonical methods directed at music researchers can be found in Windsor (2001).

For exploratory purposes, data reduction and scaling techniques, such as principal component analysis and multi-dimensional scaling (MDS), have frequently been used in music cognition studies. Examples include the semantic differential as a technique of data collection along with subsequent factor analysis introduced to German music psychology by Reinecke in the 1960s (see Böttcher & Kerner, 1978, for an overview). Principal component analysis has also been used more recently to simplify cognitive models, such as models for melodic expectation (Schellenberg, 1997).

2.2 Current models of multidimensional scaling

Multidimensional scaling has mainly been used as a means to reveal and visualise unobservable cognitive or judgemental dimensions that are of importance in a specific domain, such as tonal or harmonic relationships (Krumhansl & Kessler, 1982; Bharucha & Krumhansl, 1983), timbre similarity (Kendall & Carterette, 1991, Markuse & Schneider, 1996), or stylistic judgements (Gromko, 1993). In general, experiments start from similarity judgements that N subjects make about $P_J = \binom{n}{2}$ unordered pairs out of J auditory stimuli and aim at positioning the J auditory stimuli in a low-dimensional space with R dimensions where $R \ll P$. The R dimensions are then often interpreted as perceptual dimensions that govern human similarity perception and categorisation. Classical MDS algorithms position the judged objects in an Euclidean space where the distance, $d_{jj'}$ between the stimuli j and j' is given by

$$d_{jj'} = \left[\sum_{r=1}^{R} (x_{jr} - x_{j'r})^2 \right]^{\frac{1}{2}}$$

where x_{jr} is the coordinate of musical stimulus j on the (perceptual) dimension r. Since the Euclidean model cannot reflect differences between different sources of judgements, subjects' individual judgements are often aggregated before classical MDS solutions are computed. To accommodate individual differences between subjects, Carroll and Chang (1970) proposed the INDSCAL model where a weight w_{nr} is introduced that reflects the importance of the perceptual dimension r for subject n ($n = 1, \cdots, N$)

$$d_{jj'} = \left[\sum_{r=1}^{R} w_{nr} (x_{jr} - x_{j'r})^2 \right]^{\frac{1}{2}}$$

with $w_{nr} \geq 0$.

The INDSCAL model removes rotational invariance and makes the MDS model easier to interpret since the potentially many rotational variants of a given model do not have to be considered for interpretation any more. In turn, it generates $N \times R$ weight parameters w_{nr} for every combination of subjects and dimensions. However, these parameters are only of marginal interest if the goal of the investigation is to discover perceptual dimensions that are generalisable to a larger population of listeners. A very interesting extension of the classical MDS model was proposed by Winsberg and De Soete (1993), labelled CLASCAL. In CLASCAL models, the numerous separate parameters for every combination of subject and dimension, are replaced by so-called latent classes T which reflect types of subjects' judgemental behaviour. Since the number of latent classes is assumed to be much lower than the number of subjects N, considerably fewer weights w_{tr} for the combinations of dimensions and latent classes have to be estimated than in the INDSCAL model. The CLASCAL model by Winsberg and De Soete (1993) is given by

$$d_{jj't} = \left[\sum_{r=1}^{R} w_{tr}(x_{jr} - x_{j'r})^2 \right]^{\frac{1}{2}}$$

where w_{tr} is the weight for the latent class t with respect to dimension r. The concept of latent classes that group together subjects with similar judgemental behaviour can potentially be very useful for future research in music psychology when researchers want to explicitly allow for the possibility that subjects perceive and judge musical stimuli differently, but where a notion of *wrong* or *right* perceptions or judgements does not apply. Divergent judgements of musical or aesthetic objects in general might be caused by differences in subjects' musical backgrounds, their degree of specialisation and familiarisation with certain musical styles, or simply by their taste.[2]

The latent class approach seems to be a good compromise between the reductionist approach of just considering the mean of human judgements on the one hand, and models that require one or more parameter value for each subject (e.g. the INDSCAL model) on the other hand, yielding results that are hard to generalise. The latent class approach is also consistent with the general assumption that there are equally valid but substantially different ways to judge musical objects that, however, for a homogeneous population of listeners are limited in number. For demonstration purposes, we turn to timbre perception as the study area. Here, a number of publications make use of an extended CLASCAL model as the statistical method. The extended model is given by

$$d_{jj't} = \left[\sum_{r=1}^{R} w_{tr}(x_{jr} - x_{j'r})^2 + \nu_t(s_j + s_{j'}) \right]^{\frac{1}{2}}$$

where s_j and $s_{j'}$ are coordinates on dimensions that are specific to the objects j and j', i.e. not shared with other objects. The parameter ν_t is the weight a latent class gives to the whole set of specific dimensions. It reflects how much a perceptual strategy uses common and general perceptual dimensions to compare musical objects or how much it attends to the specificities of each musical stimulus. McAdams, Winsberg, Donnadieu, De Soete, and Krimphoff (1995) applied this extended CLASCAL model to 153 pairwise comparisons of 18 instrument timbres as obtained from judgments by 88 subjects with different degrees of musical training. The authors found five different latent classes of judgmental behaviour and obtained a three-dimensional model with instrument-specific dimensions. These five latent classes were only very partially explained by the amount of musical training

[2] It is not surprising that latent class models have been quite popular in food science where subjects' judgements in sensometric experiments are usually not regarded as being right or wrong but simply as different from one another (Sahmer, Vigneau, & Qannari, 2006)

subjects had received. Instead it seems likely, that the latent classes reflect different judgemental behaviour in terms of the perceptual dimensions attended to. McAdams et al. (1995) were able to correlate the three dimensions of their model very closely with three different acoustic parameters previously investigated by Krimphoff, McAdams, and Winsberg (1994), namely log-attack time, the spectral centroid, and spectral flux. They also suggested some distinguishing characteristics that could be aligned to the specificities of each instrument, for example the *hollow timbre* of the clarinet or the *very sharp, pinched offset with clunk* of the harpsichord. These examples show the point of employing an MDS model to specificities: The sharp pinched sound of a harpsichord is only at a very abstract level comparable to the hollow timbre of a clarinet, but when the clunk is present in a sound it can be a very important factor for timbre perception. A more recent study by Caclin, McAdams, Smith, and Winsberg (2005), employing the recent CONSCAL model, confirmed the close relations between perceptual dimensions and attack time, spectral centroid, and spectral flux, and also discovered the significant role of the attenuation of even harmonics. The work on scaling techniques and their application to timbre perception is on-going (see Burgoyne & McAdams, 2007), and it might generate some interesting results on individual perceptual differences between different groups of subjects along with models describing the different perceptual strategies.

2.3 Classification and regression trees

For a long time classification and regression trees have been part of the tools and techniques for general data mining and machine learning. However, there have been only a few studies in music psychology that make use of the inherent advantages of these models as we shall see below. Although various software packages for statistical computing (SPSS, SAS, Statistica, R) have implemented several different but related models and algorithms (e.g. CHAID, ID3, C4.5, CART), we just take two studies as examples that both make use of the CART algorithm as implemented in R.

Classification and regression trees partition observations into categories along a dependent variable y by using the information in the set of independent variables X. The goal is to arrive at a robust model that enables us to predict the value of y given the values for X in a future case. Tree models belong to the area of supervised learning which means that they learn from complete datasets that contain a sufficiently large number of cases with given values for y and X. The models are visualised in a very straight-forward manner as tree structures where nodes represent subsets of the original dataset and branches lead from one subset of the data to two subsets at the next lower level. The first node contains the original dataset. It is called *root*, and the algorithm starts here by recursively partitioning the dataset into smaller and smaller subsets. Figure 1 gives an example of the visualisation of a tree model. The example is taken from a recent study on melodic accent perception (?, ?) where tree models are employed to predict the perceptual accent strengths of individual notes in a melody on the basis of a large set of rules, mainly derived from Gestalt laws.

For the construction of a classification or regression tree one has to decide on three questions (see Breiman et al., 1984):

1. When and how should a node be split into two subnodes?
2. When is a node an end node that should not be split any further? In other words, how should the size and complexity of the tree be determined?
3. How should the end nodes (*leaves*) be labelled, i.e. how does the tree make predictions?

The first two questions rely on measures of the classification or regression accuracy of the tree as a whole. They decide on the question whether or not the classification accuracy of the dataset as a whole is better if a node is split into two subnodes. For classification trees where the dependent variable y is categorical, the Gini-Index is a widely used measure. It is defined as

$$I(t) = \sum_{j \neq i} p(i|t)p(j|t)$$

where $I(t)$ denotes the impurity of the observations in a node and $p(i|t)$ is the probability for a random observation t to be assigned to class i. By recursively splitting the tree into subbranches, the algorithm seeks to minimise the Gini-Index. For interval-scale-dependent variables, the so-called *ANOVA criterion* is generally employed for splitting a node. This criterion is defined as $SS_K - (SS_L + SS_R)$ where $SS_K = \sum (y_i - \overline{y})^2$ and SS_L and SS_R are the sum of the squared differences in the left and right subnode. The ANOVA criterion is sought to be maximised in such a way that at each node with each independent variable x the partitioning algorithm tests whether the accuracy criterion can be improved by splitting the data into two subnodes according to all potential values of x. Where the improvement in the criterion is largest, the value of x is then chosen as the splitting or decision value. The recursive partitioning into smaller and smaller subnodes is carried out until the endnodes (leaves) only contain a minimal number of observations. However, in order to generate a stable tree that is not overfitted on the learning dataset, the full tree has to be *pruned* subsequently. In other words, only true relations between dependent and independent variables should be reflected in the tree model. This is achieved by using the so-called *cross-validation* method. Cross-validation uses

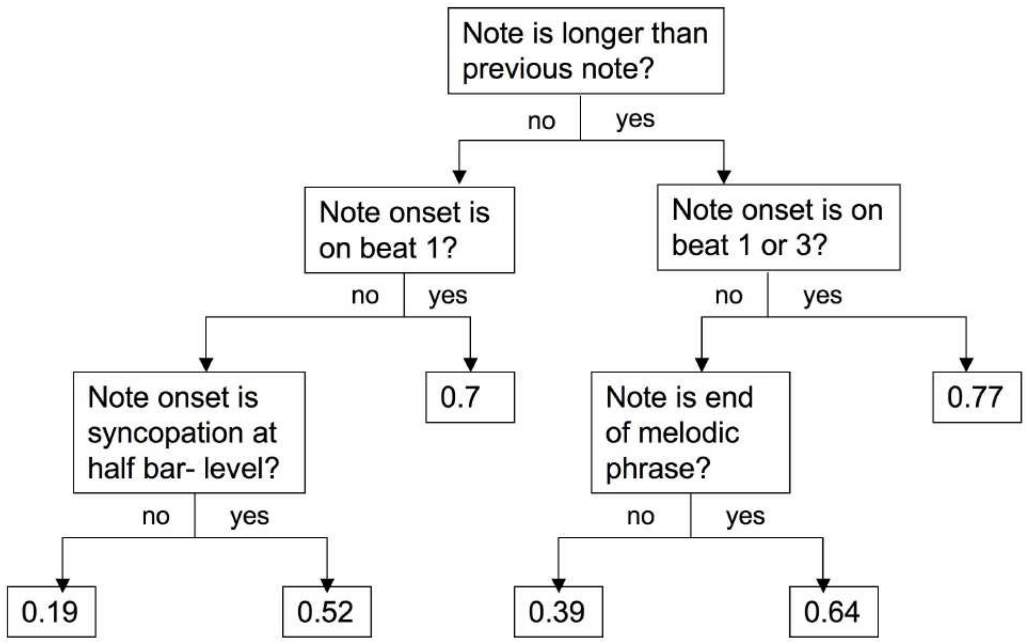

Figure 1: Regression tree model from Müllensiefen et al. (under review). The graph reads from top to bottom. Observations (here: notes of a melody) are recursively partitioned into smaller subsets by binary splits at each node. All observations satisfying the splitting criterion (i.e. answering 'yes' to question in box) are gathered together in the child node to the right and all observations not satisfying the condition form the subset represented by the child node to the left. The end nodes at the bottom of the graph contain prediction values (here: perceptual accent strength on a scale from 0 to 1) for all observations classified by the sequence of binary conditions in the same way, from the root to the end node.

the information from a larger portion of the data to build a tree model which then predicts the observations of a smaller data subset. The classification error of the cross-validated tree model is usually getting smaller as the tree model increases in complexity (i.e. it has more nodes). However, from a certain degree of complexity, the cross-validation error arrives at a *plateau* or even increases with increasing complexity. Therefore, Therneau and Atkinson (1997) suggest a good balance between partitioning accuracy and tree complexity so that tree stability is reached when the cross-validation error has reached its plateau.

As an example, Müllensiefen and Hennig (2006) use classification and regression trees among other techniques from the data mining repertory (including random forests, linear and ordinal regression, and k-nearest neighbour) to explain the participants' responses in a music memory task. The task consisted of spotting differences between a target melody in its musical context and an isolated comparison melody, similar or identical to the context melody. The study aimed at identifying the factors that determine recognition memory for new melodies and tunes (see also exp. 2 from Müllensiefen, 2004). As the most important predictors for explaining these memory recognition judgements, the overall melodic similarity, and the similarity of the melodic accent structures, as well as the subjects' musical activity were identified.

Another example for the application of classification and regression trees is a study by Kopiez, Weihs, Ligges, and Lee (2006) where the authors try to predict performance in a sight reading task. The predictor (independent) variables in this study include general and elementary cognitive skills as measured by standard psychological tests, as well as practice-related skills and the amount of time invested in musical activities. For classifying 52 subjects into good and bad sight readers, linear discriminant analysis gave slightly better results than a classification tree from the CART family. Among the most important factors to predict high achievements in the sight reading task were the subjects' speed at playing trills, their mental speed as measured by a number connection test, and the time invested in sight reading before the age of 15.

Taken together, both application examples show that classification and regression trees are only one out of a larger number of statistical techniques that may be used as classification or prediction models, although tree models might not necessarily deliver the most accurate prediction results. Nonetheless, they have a few particular advantages that seem to fit circumstances well in many music psychology experiments. Among these advantages are:

1. If many independent variables can be assumed to influence the dependent variable under study, selection mechanisms become quite important to identify the variables with most explanatory power. Two mechanisms already built into tree models serve this need for variable selection., which are, first,

recursive partitioning by the variables that provide the largest increase in the accuracy criterion, and second, tree pruning as such.

2. In data sets from music psychology experiments, predictor variables (e.g. variables on musical background) often have missing values for some observations (subjects), and the concept of surrogate predictor variables (not explained here due to space limitations) copes easily with these cases.

3. In many cases, higher-order interactions between several predictor variables cannot be ruled out from the model a priori. Tree models represent these interaction terms very effectively. In fact, tree models might be regarded as models of variable interactions only, mainly ignoring additive effects that are better modelled by linear models.

4. Non-linear relationships between the dependent and one or more independent variables can be accommodated for in tree models, while linear models, by definition, are rather poor at modelling non-linear relationships.

5. Tree models make only few assumptions regarding the distribution of the data which is in contrast to many linear models that rely, for example, on the normal distribution of the residuals.

6. The option to visualise tree models as graphs that even non-scientists can understand by intuition, yields the potential to popularise and communicate research results beyond a circle of experts.

It quite likely that these advantages make tree models an attractive tool for music psychologists in the future, especially in comparison with linear models that are currently by far more popular in this study area.

2.4 Functional data analysis

Given that music is a time-dependent domain and music cognition presumably evolves in time as well, it is surprising how few studies in the field of music cognition make explicit use of time related information in their statistical analysis (for a notable exception see the study of continuous emotional responses during music listening based on time series analysis by Schubert and Dunsmuir (1999)).

Functional data analysis (FDA) is a relatively new statistical concept that is particularly well-suited to represent and analyse how musical parameters or observed human reactions to music evolve over time. A functional datum is not a single observation but a set of measurements along the time axis that can be regarded as a single entity (see Levitin et al., 2007). This means that in FDA, observations are curves of random functions and not values of random variables.

Even though data are only sampled at discrete interval steps along a continuum, the aim of FDA is to express the variation in the variable of interest as a function of a continuous variable, most commonly time. Thus, FDA provides the ability to quantify the dynamics of the variable under study.

A typical FDA runs through a series of steps which we here describe very briefly following the detailed explanations in Ramsay and Silverman (2005); Levitin et al. (2007).

1. **Data gathering** The raw values of the dependent variable y_j are recorded at j discrete points of the continuous variable t_j with $j = 1, \cdot, n$. For applications in music research the continuous variable is most commonly time. A single functional observation comprises n tuples (t_j, y_j). If replications of the functional process are recorded (e.g. by testing several subjects), the index i is used to refer to the different replications, and tuples are double indexed (t_{ij}, y_{ij}).

2. **Smoothing and interpolation** FDA assumes that raw data y_j are generated by a latent underlying process that is best represented by a smooth and continuous curve. The underlying process is denoted by $x(t_j)$ and its relation to the observable data is described by adding a noise or error term ϵ_j:

$$y_j = x(t_j) + \epsilon_j$$

   One of the differences between FDA and many other statistical methods is that FDA does not assume that the error term is independently distributed over observations, nor that it has a mean of zero and a constant variance of σ^2. In fact, for many biological and psychological processes (e.g. blood pressure, heart rate, emotional arousal, strength of subjective mood), errors for close-in-time observations might be correlated, although they might differ systematically at distant points in time. $x(t_j)$ is obtained by fitting a set of so-called basis functions ϕ_k to the raw data. $x(t)$ is then represented by the sum of the basis functions ϕ_k weighted by their corresponding coefficients c_k:

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

   There are many candidates for suitable basis functions, including the Fourier series, wavelets, local polynomial regression, and B-splines (e.g. Schumaker, 1981). The choice of a particular type of function for a given dataset depends on whether a function is good at reflecting the noise generating process so that a low number k for functions and coefficients is sufficient to represent these data.

3. **Aligning the smooth data of the i replications** For this processing step, data from the performance of the same piece of music by different musicians and at different tempos can be taken. If differences between replications (e.g. subjects) are not of interest, data can be averaged at this stage.

4. **Displaying the main dynamic characteristics of the data** For this processing step, smoothed data are differentiated. In fact, one main reason for smoothing the data in FDA is to use differential equations as models. A very popular way to display the evolution of the underlying process in time is to plot the first and second derivatives against each other to make the evolution of slowness or acceleration of the variables under study visible. This yields the so-called *phase-plane plot*.

5. **Modelling the aligned data** Within the tool set of functional analysis, there are many adaptions of standard statistical techniques available. These include functional principal component analysis and functional linear modelling. Just to give a simple example of a functional variant, we consider the simple case of a functional linear model where the dependent variable is functional but independent variables are not. The functional linear model describing all replications i is defined as

$$y_i(t) = \sum_{j=1}^{J} \beta_j(t) x_{ij} + \epsilon_i(t)$$

   where the regression coefficients $\beta_j(t)$ are functions of time. For modelling the aligned data, linear models where functional dependent variables are predicted by a set of functional independent variables are also possible, and the reader is referred to Ramsay and Silverman (2005).

Up to now only a few studies have been published in music psychology using FDA to cover very different topics. Vines, Nuzzo, and Levitin (2005) and Vines, Krumhansl, Wanderley, and Levitin (2006) use functional data analysis to study perceived tension during music perception. In their experiment, musical tension is measured continuously in musically trained participants who operate a slider device when three different conditions are given, a) when listening to an excerpt of Stravinsky's second piece for clarinet solo, b) when listening and watching a clarinettist play the piece, and c) when only watching the video. The data of the continuous slider are sampled at high frequency, and the resulting curve is fitted by a large number of 6th order B-splines. As one result of the functional data analysis, Vines et al. present phase-plane plots (see Figure 2).

By analogy with well-known concepts in physics, they analyse the participants behaviour in terms of *emotional kinetic* and *potential energy* and show that the

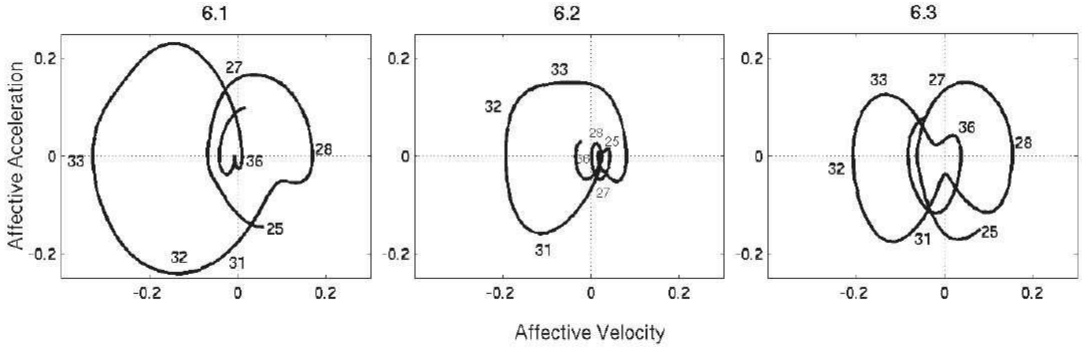

Figure 2: Phase-plane plots from Vines et al. (2005) illustrating the evolution of affective tension over time. *Affective Velocity*, x axis, is plotted against *Affective Acceleration*, y axis. The integers denote different points in time of the piece as measured in seconds from the beginning. Panel 6.1 shows affective tension from the auditory-only condition, panel 6.2 gives tension during visual presentation, and panel 6.3 shows perceived tension from the audio-visual condition.

evolving emotional experience over time relates roughly to specific events in the compositional structure of the piece. They found that the additional visual information served different purposes at different points in listening. For some passages it reduced the amount of perceived tension while at other instances it increased tension experience. For mere visual presentation they recorded a much lower flow of affective energy. However, the authors found that additional visual information helped subjects understand the performer's phrasing and to anticipate changes in emotional content. Tension perception was clearly influenced by the phase-advanced visual information in the audio-visual condition. Vines et al. (2006) conclude that there might be emergent perceptual qualities when music is both heard and seen.

Almansa and Delicado (2009 (in press)) apply FDA to a quite different set of data. Instead of investigating perception, they look at tempo variations in 28 different performances of Robert Schumann's Träumerei. The data were collected by Repp (1992) and consist of a 28X253 matrix with rows corresponding to performances and columns corresponding to the 253 crotchets of the piece. The measurements reflect tone duration in milliseconds for each crotchet. Almansa and Delicado (2009, in press) use local polynomial regression to smooth the data and then align the data of the 28 performances by score time. The authors then perform a functional principal component analysis (fPCA) on the smoothed tempo data and find a number of meaningful components that lend themselves readily to musical interpretations. Among the components that explain most of the variance in the data are size, ritardando, contrast, as well as a period-wise component. The *size* component reflects the global tempo, the *ritardando* component describes the differences between the global tempo and the final ritardando, while the *contrast* component compares the faster performances of phrases A and B in opposition to a stronger slowing down at the end of phrase B and the final fermata. The *period-wise* component describes the generally slower tempo in the middle of phrases A, B, and A' in contrast to faster tempo at the end of the phrases. As a last step, Almansa and Delicado (2009 (in press)) apply a hierarchical cluster analysis to the performance data after analysis with fPCA. They arrive at a clustering solution with four clusters comprising five to eight performances that can be considered similar to each other in terms of the functional principal components. Accordingly, the approach to combine fPCA with subsequent clustering might be a suitable general and robust method to compare and classify the structure of musical performances despite superficial differences. If, for example, global tempo is not of interest, the first component could be left out and classification could be based on the remaining components reflecting more subtle usages of musical tempo. This approach is general enough that, apart from tempo, other parameters like performance, loudness, or timbre register could be used for comparison and modelled in a functional way.

In general, FDA opens up a range of perspectives for music analysis, since music is a time-dependent phenomenon and many musical parameters as well as human reactions to music can be assumed to change continuously during listening. However, an important requirement is the availability of large sets of data. This makes FDA particularly useful for the analysis of audio performance data sampled at high rates, and also of neuronal data recorded with techniques of high temporal resolution such as EEG, ERPs, and MEG.

2.5 Bayesian models of music perception

Bayesian reasoning and probabilistic models have received a lot of attention in cognitive psychology for quite a while now (e.g. Chater et al., 2006; Chater & Manning, 2006). But for some unclear reason, application of these ideas in music research has been rather sporadic, even though scholars since Meyer (1957) have noted that many concepts in music might lend themselves naturally to a formulation in probability theory. Bayesian modelling constitutes a framework for reasoning with uncertainty. According to Chater and Oaksford (2007) this framework can be applied to two different realms of psychological research. First, Bayesian models can model scientific datasets that stem from psychological experiments or observations. Here, and in contrast to traditional statistical techniques from the Nyman-Pearson or the Fisher school, Bayesian models quantify the researcher's prior beliefs, assumptions, and uncertain knowledge, and take this external information into account when comparing different models of the experimental data. Bayesian models therefore represent an alternative to the hypothesis testing approach that has been dominating empirical psychology. Second, Bayesian models are not only an alternative rational way to experimentally analyse data, they can also serve as models of cognition whenever the human mind is regarded as a Bayesian reasoning system. This way, prior beliefs and knowledge of participants can be modelled along with stimulus data subjects might be presented with during a perception experiment. The observed experimental responses can then be modelled via Bayes' theorem. Bayes' theorem can be deduced from the axioms of basic probability theory and is expressed in terms of conditional probabilities. The notation of the conditional probability $P(A|B)$ denotes the probability of A being true, given that B is true. Bayes' theorem is defined as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

To give an example of how this equation might serve as a cognitive model in music research, we might imagine an experimental task like the one described in Lippens, Martens, De Mulder, and Tzanetakis (2004) where subjects are asked to name the musical genres after hearing excerpts of pop music pieces. Let us denote genre labels by A and perceived musical characteristics by B. According to the Bayesian model subjects would give the musical genres a label with the highest probability given the musical characteristics just perceived, i.e. a label with highest *posterior probability* maximises $P(A|B)$. According to the Bayesian model, subjects make use of their knowledge about the conditional probability $P(B|A)$ resulting from the musical characteristics B in music from genre A (e.g. distorted guitars in heavy metal songs). Subjects would also make use of their prior belief $P(A)$ of how likely it is that songs from genre A are actually occurring as test stimuli in a psychological experiment. $P(A)$ is also called the *prior probability* or just *prior*. Finally, $P(B)$, which is the general probability of the musical characteristics (e.g. distorted guitars) just heard, is taken into account.

While Bayesian modelling of musical genre perception is an interesting thought experiment, and an empirical study still has to be carried out, a highly interesting application of Bayesian modelling is provided by Sadakata and colleagues (2006). The authors provide a meta-study of human rhythm perception and production. The starting point is the apparent psychological asymmetry between performance and perception of musical rhythms. While for the perceptual dimension, duration changes between subsequent notes tend to be emphasised, the inequality in the durations of subsequent notes tends to be assimilated in rhythm production. Sadakata et al. (2006) use existing data from four different studies on rhythm production and perception. Their goal is to model the perceptual data on the basis of the production data using Bayes theorem. In particular, they model the posterior probability $P(c|t)$ that denotes the probability that subjects perceive the rhythmic class of two successive notes c given that they are aurally presented with time ratio t, i.e. the ratio of the inter-onset intervals of two consecutive notes as measured in milliseconds. The rhythmic class c essentially represents how long the first note is in comparison to the second one (e.g. equal duration values or punctuated rhythms). Inserted into Bayes' equation their model looks like this:

$$P(c|t) = \frac{P(t|c)P(c)}{P(t)}$$

As estimates for the conditional probability $P(t|c)$, meaning that a time ratio t in milliseconds is produced given a rhythm class c, Sadakata et al. (2006) use the data from their rhythm production experiments. For estimating the priors, they use three different sources: a uniform distribution where priors for all rhythm classes c are equal, a theoretical rhythm complexity model derived from the Farey tree (Peper et al., 1995), and a distribution derived from frequency counts of rhythms in three different music corpora. They also looked at how perceived rhythm classes were predicted just by the conditional probability $P(t|c)$ alone, i.e. without using any priors. Just for the sake of comparison they also produced a model where optimal priors were obtained by fitting the production data to the perceptual data. As expected, their results showed that the no-prior model always gave the worst predictions, while the optimal priors model was in all cases superior to all other models. Apart from these trivial results Sadakata, Desain, and Honing (2006) found that the priors that made use of information from the Farey tree or from the music corpora generally outperformed the model with uniform priors. This result neatly shows how the incorporation of quantified prior musical knowledge into models of music perception can greatly enhance model accuracy. This point is also made by Müllensiefen, Wiggins, and Lewis (2008) who endorse the concept of *corpus-based* musicology, i.e. the idea of using musical knowledge as extracted from large music corpora to enhance the predictive power of models of music perception. The meta-study by Sadakata et al. (2006) implicitly makes another important point in that it shows that even in psychological studies that sometimes use only vaguely music-like stimuli, participants seem to relate these stimuli to prior musical experience and perceive them on these grounds. In general, Bayesian analysis provides a useful and straightforward framework for incorporating beliefs, assumptions, as well as prior musical knowledge as possessed by either the experimenter or the experimental participants into the statistical analysis of experimental results. This means also that the access to relevant and meaningful prior knowledge is all-important. Indeed, the biggest advantages of Bayesian modelling come into play when musical knowledge is quantifiable and is also assumed to influence reactions and decisions in an experimental task.

3 Music Modelling

The distinction between research in cognitive music psychology and what we call *music modelling* is admittedly not very clear-cut and sometimes simply a matter of perspective. While the studies reviewed in the last section are primarily interested in mental processes connected to music perception and cognition by observing human behaviour that reflects these mental processes, the studies that we now look at rather try to describe the structure of music itself statistically. Of course, descriptions and representations of musical structure are always a result of human cognition, but studies of music modelling tend to be less interested in the nature of the underlying psychological processes that generate musical structures than in the structures themselves.

Modelling music data has become increasingly popular in recent times, due largely to the increasing amount of music that is digitally available in a symbolic encoding format (e.g. MIDI, EsAC, kern, or other codes that encode notes as the basic musical events). While statistical approaches for describing the compositional structure of music have been present since the 1950s (e.g. Meyer, 1957; Moles, 1958; Steinbeck, 1982; Fucks, 1962; Fucks & Lauter, 1965; Fucks, 1968), the number and diversity of statistical approaches for modelling structural features of music have rocketed over the last decade.

3.1 Bayesian models of musical structure

Also for describing musical structures, Bayesian modelling has become quite popular (e.g., Temperley, 2004, 2006, 2007; Rhodes et al., 2007). As an example, we take Temperley's description of a Bayesian model that determines the key of a given piece or segment of music (Temperley, 2004). At its core we find Bayes' rule that calculates the probability of a musical feature or structure (here: the key) given an empirical music surface (here: the frequencies of pitch classes in a musical segment).

$$P(\text{key structure} \mid \text{surface}) \propto \prod_{segment} \left(M \prod_{p} K_{pc} \prod_{\sim p} (1 - K_{pc}) \right)$$

Here, M is a modulation score that penalises the change in key from one segment to the next and K_{pc} stands for the key-profile values of the pitch-classes present (p) in a segment which are multiplied by the product of $(1 - K_{pc})$ for all pitch-classes not present in this segment ($\sim p$). Thus, the probability computation for musical keys, given the pitch classes of a musical piece, is based on the relative frequency with which the twelve scale degrees appear in a key as well as on the probability of a subsequent segment of the piece being in the same key as the previous segment. Comparable to the testing of different priors as in the study by Sadakata et al. (2006), Temperley uses pitch-class profiles derived from different sources (music collections or corpora). Tested against other key finding models like his own non-Bayesian model (Temperley, 2001) or the Krumhansl-Schmuckler algorithm (Krumhansl, 1990), his Bayesian model achieves about equal success rates, although, as Temperley openly admits, most of the core features of the model may be formulated without Bayesian terminology and notation. Reviewing his recent book (Pearce et al., 2007) on probabilistic models and music (Temperley, 2007), it seems that Temperley has not yet made full use of the potential that the Bayesian approach can offer for modelling musical structures. It appears that Bayesian modelling in this sense is largely concerned with precisely quantifying rule-based systems of musical analysis. Therefore, it remains to be seen what the original Bayesian contribution to these kinds of musical models will be, considering that frequency counts on musical elements have been successfully used as predictors in non-Bayesian models before (e.g. Eerola et al., 2002; Costa et al., 2004).

3.2 n-gram models of musical structure

Another recent, prominent trend in music modelling is the use of Markov-chains or n-gram modelling. Here, the basic musical units are longer sequence structures, instead of single events such as pitches, intervals, or durations. This approach builds on the basic assumption that music is principally produced and perceived as a time-ordered set of events, be it tones of a melody or harmonies in a polyphonic piece. The notion that music can be explained, taught, and analysed as formulae has been around for several hundred years in music theory, but only due to the recent availability of large electronic corpora can these hypotheses be empirically tested.

A sophisticated example of the n-gram approach is the work of Pearce and Wiggins (e.g. 2004, 2006) which is concerned with melodic n-grams. Their research hypothesis is that many aspects of musical expectation are acquired through spontaneous induction of sequential regularities in the music we are exposed to. Consequently, they name their model the *Information Dynamics of Music* model, or short the *IDyoM* model, and define it as a model of sequences e_i composed of symbols drawn from an alphabet $\mathcal{E}$. The model estimates the conditional probability of an element at index i in the sequence, given the preceding elements in the sequence: $p(e_i|e_1^{i-1})$. Given such a model, the degree to which an event appearing in a melody is unexpected can be defined as the *information content* (MacKay, 2003), $h(e_i|e_1^{i-1})$, of the event, given the context:

$$h(e_i|e_1^{i-1}) = \log_2 \frac{1}{p(e_i|e_1^{i-1})}.$$

The information content can be interpreted as the contextual unexpectedness or surprise associated with the event. The contextual uncertainty of the model's expectations in a given melodic context can be defined as the *entropy* (or average information content) of the predictive context itself:

$$H(e_1^{i-1}) = \sum_{e_i \in \mathcal{E}} p(e_i|e_1^{i-1})h(e_i|e_1^{i-1}).$$

Just as in the Bayesian model reviewed above, this model builds on counts of occurrences of melodic n-grams in a collection of melodies. For a given melodic context, the model returns the continuation of chain of *n-1* notes with maximum likelihood as a prediction for the given context. Although the idea of recording the frequencies of melodic n-grams in a corpus and returning the like for any given context sounds quite straight forward, the details of their modelling are quite intricate and consist of techniques adopted from statistical language learning and data compression, which are outside the scope of what can be explained here in detail. Pearce and Wiggins (2004) examine several model parameters, first, the model type, where they find best results by employing a combination of a long-term model trained on large corpora of melodies (e.g. the Essen folk song collection, church hymns, ballads), and a short-term model trained exclusively on the melody currently being predicted. Second, they place emphasis on the treatment of novel n-grams when encountered in a prediction context for which no frequency count yet exists in the model. Here, smoothing and escape heuristics are explored from automatic speech processing. As a third and important model parameter, they vary the upper order bound (maximal length) of the n-grams considered. In combination with other parameters a variant yields optimal model performance where the order of the n-gram is unbound but the relative weight of a n-gram is adjusted according to its length, for the averaged prediction gives more weight to longer n-grams. A fourth and decisive set of model parameters is the type of abstraction or transformation to be applied to the raw melody data, where they determine a handful of musically meaningful *view points* (i.e. musical dimensions or parameters such as raw pitches, pitch intervals, note durations) through a standard variable selection process (step-wise variable elimination).

As a result from this search through the parameter space, Pearce and Wiggins (2006) are able to predict experimental data from three previous studies (Cuddy & Lunney, 1995; ?, ?, ?) where expectation of melodic continuation was determined from behavioural experiments with human listeners. Their model performs well on all three data sets and outperforms a competing model proposed by Narmour (1990) based on principles of Gestalt psychology (implemented by Schellenberg, 1997). Pearce and Wiggins' reasoning (2006) regarding model selection is a notable point of their approach and can serve as a guideline for other studies concerned with the comparison of models for music cognition. As a criterion for model selection, they do not only consider data fit, but also scope (which is the model's failure to predict random data), and simplicity (which is the number of prior assumptions and principles the model builds upon).

Strong evidence in favour of the *IDyoM* model as a general model of melody perception comes from a number of recent studies where the model was not used for its original purpose (i.e. to predict melodic expectation), but was applied to automatically segment full melodies into melodic phrases. The underlying rationale is that perceptual groups are associated with points of closure where the ongoing cognitive process of expectation is disrupted either because the melodic context fails to stimulate strong expectations for any particular continuation or because the actual continuation is unexpected. For predicting the manual phrase segmentations of 1705 folk songs from the Essen collection which were annotated by expert folk song collectors, the *IDyoM* model performed at a comparable level of accuracy to a couple of other computational models. The surprising result of this study was the fact that *IDyoM* as a model of melodic expectation did almost as well as existing models that were specifically designed for segmenting melodies (e.g. Grouper, Temperley, 2001, and LBDM, Cambouropoulos, 2001) and make use of high-level music theoretical knowledge about melody segmentation (see Pearce et al., 2008). The fact that the model generates acceptable results outside its original application domain let the authors hypothesise that mechanisms of statistical learning which are the core of the *IDyoM* model, actually represent the underlying processes of melody perception as well as the acquisition of musical knowledge.

4 Conclusion

This update intends to highlight some of the more interesting statistical approaches as employed in recent music psychology research and aims at motivating music psychologists to explore the analytical and epistemological possibilities that these new techniques provide. Naturally, a short overview like this can only be far from complete, both in terms of depth (of the mathematical technicalities and the design of studies reviewed) as well as in terms of the range of new techniques covered. Nonetheless, we hope that this paper serves as an overview of current trends so that music psychologists might obtain an idea of where empirical methodology is heading. Some final remarks have to be made on the available software that allows to compute analyses using the methods described. Only if software is available and accessible as well as easy to handle, new statistical techniques have a chance of becoming a popular research tool and will possibly be integrated into the canon of methods. While in the past, music psychologists were often dependent on a few specialised companies to integrate a new technique into commercial software packages such as SPSS, SAS, or Statistica with graphical user interfaces (GUIs), the success of powerful and yet high-level programming languages that are specially designed for data analysis or, at least, have large libraries designed for scientific statistical computing, opens a new range of possibilities. The most important of these very high-level data analysis environments are Weka[3] (which includes a GUI), Matlab[4] (*Statistics Toolbox*), R[5], S-Plus[6], and Python[7](*SciPy* and *StatPy* packages). Researchers and software developers around the globe constantly contribute codes for new statistical analysis to these environments that are then available to music psychologists who, in general, are not keen to implement new statistical procedures from scratch themselves. Apart from Matlab and S-Plus, which require an expensive license for the basic system and commercial toolboxes, the usage of the other aforementioned data analysis environments is free.

Regarding the statistical techniques covered in this paper, general multidimensional scaling packages, including INDSCAL, are implemented in most commercial as well as free software programmes. Unfortunately, the discussed CLASCAL model is not yet available in any larger environment. In contrast, classification and regression trees (often called decision trees) have implementations in most environments. The studies cited above made use of the R-package CART. Functional data analysis packages are maintained for Matlab, R, and S-Plus. For basic reasoning with Bayes' theorem, no special software is required. But there is a wealth of advanced Bayesian methods available that we did not cover here, since for almost all software environments more or less comprehensive Bayesian packages or libraries are obtainable. Finally, the *IDyoM* model is unfortunately not publicly available. But many programmes provide the basic tools for sequence-based methods often employed in computational linguistics (e.g. package *tm* in R, *Clementine* in SPSS or *KEA* in Weka).

Acknowledgements

Daniel Müllensiefen is supported by EPSRC grant EP/D038855/1. We would like to thank Alex McLean and Klaus Frieler for valuable comments on earlier versions of the manuscript.

[3]http://www.cs.waikato.ac.nz/ml/weka/
[4]http://www.mathworks.com/
[5]http://www.r-project.org/
[6]http://www.insightful.com/
[7]http://www.python.org/

References

Almansa, J., & Delicado, P. (2009 (in press)). Analysing musical performance through functional data analysis: Rhythmic structure in schumann's träumerei. *Connection Science*.

Beran, J. (2004). *Statistics in musicology*. Boca Raton: Chapman Hall.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, *5*, 1-30.

Bharucha, J. J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, *13*, 63-102.

Böttcher, H. F., & Kerner, U. (1978). *Methoden in der musikpsychologie*. Leipzig: C. F. Peters.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Belmont (CA): Wadsworth.

Burgoyne, J., & McAdams, S. (2007). Non-linear scaling techniques for uncovering the perceptual dimensions of timbre. In *In proceedings of the international computer music conference (icmc 2007)* (Vol. 1, p. 73-76).

Caclin, A., McAdams, S., Smith, B., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, *118*(1), 471-482.

Cambouropoulos, E. (2001). The local boundary detection model (lbdm) and its application in the studt of expressive timing. In *Proceedings of the international computer music conference* (p. 17-22). San Francisco: ICMA.

Carroll, J., & Chang, J. (1970, 283-319). Analysis of individual difference sin multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika*, *52*, 35.

Chater, N., & Manning, C. (2006). Probabilistic models of language processing and acquisition. *TRENDS in Cognitive Sciences*, *10*(7), 335-344.

Chater, N., & Oaksford, M. (2007). *Bayesian rationality. the probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Chater, N., Tenenbaum, J., & Yuille, A. (2006). Probabilistic models of cognition: conceptual foundations. *TRENDS in Cognitive Sciences*, *10*(7), 287-291.

Costa, M., Fine, P., & Ricci Bitti, P. E. (2004, September). Interval distribution, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception*, *22*(1), 1-14.

Cuddy, L. L., & Lunney, C. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics*, *57*, 451-462.

Desain, P., & Honing, H. (1992). *Music, mind and machine, studies in com-*

puter music, music cognition and artificial intelligence. Amsterdam: Thesis Publishers.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2002, March). Statistical features and perceived similarity of folk melodies. *Music Perception*, *18*(3), 275-296.

Flexer, A. (2006). Statistical evaluation of music information retrieval experiments. *Journal of New Music Research*, *35*(2), 113-120.

Fucks, W. (1962). Mathematical analysis of formal structure of music. *IEEE Transactions Information Theory*, *8*(5), 225-228.

Fucks, W. (1968). *Nach allen regeln der kunst*. Stuttgart: DVA.

Fucks, W., & Lauter, J. (1965). *Exaktwissenschaftliche musikanalyse*. Opladen: Westdeutscher Verlag.

Gromko, J. E. (1993). Perceptual differences between expert and novice music listeners at multidimensional scaling analysis. *Psychology of Music*, *21*, 34-47.

Kendall, R. A., & Carterette, E. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, *8*, 360-404.

Kopiez, R., Weihs, C., Ligges, U., & Lee, J. I. (2006, October). Classification of high and low achievers in a music sight-reading task. *Psychology of Music*, *34*(1), 5-26.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. ii: Analyses acoustique et quantification psychophysique. *Journal de Physique*, *4*(C5), 625-628.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Krumhansl, C. L., & Kessler, E. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*, 334-368.

Levitin, D. J., Nuzzo, R. L., Vines, B. W., & Ramsay, J. O. (2007). Introduction to functional data analysis. *Canadian Psychology*, *48*(3), 135-155.

Lippens, S., Martens, J.-P., De Mulder, T., & Tzanetakis, G. (2004). A comparison of human and automatic musical genre classification. In *Ieee international conference on acoustics, speech, and signal processing (icassp)* (Vol. 4, p. 223-236).

MacKay, D. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.

Markuse, B., & Schneider, A. (1996). Ähnlichkeit, nähe, distanz: zur anwendung multidimensionaler skalierung in musik-wissenschaftlichen untersuchungen. *Systematische Musikwissenschaft/ Systematic Musicology/Musicologie systématique*, *4*, 53-89.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual sclaing of synthesized musical timbres: ommon dimensions, specificities, and latent subject classes. *Psychological Research*, *58*, 177-192.

Meyer, L. B. (1957). Meaning in music and information theory. *Journal of Aesthetics and Art Criticism*, *15*, 412-424.

Moles, A. (1958). *Théorie de l'information et perception estétique.* Paris: Flammarion.

Müllensiefen, D. (2004). *Variabilität und konstanz von melodien in der erinnerung. ein beitrag zur musikpsychologischen gedächtnisforschung. [variability and constancy of remembered melodies: A music-psychological investigation of memory].* Unpublished doctoral dissertation, University of Hamburg.

Müllensiefen, D., & Hennig, C. (2006). Modeling memory for melodies. In M. Siliopoulou, R. Kruse, C. Borgelt, A. Nürnberger, & W. Gaul (Eds.), *From data and information analysis to knowledge engineering* (p. 732-739). Berlin: Springer.

Müllensiefen, D., Wiggins, G., & Lewis, D. (2008). High-level feature descriptors and corpus-based musicology: Techniques for modelling music cognition. In A. Schneider (Ed.), *Systematic and comparative musicology: Concepts, methods, findings* (Vol. 24, p. 133-155). Frankfurt: Peter Lang.

Narmour, E. (1990). *he analysis and cognition of basic melodic structures: The implication-realization model.* Chicago: University of Chicago Press.

Orio, N. (2006, November). Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, *1*(1), 1-90.

Pearce, M. T., Müllensiefen, D., Rhodes, C., & Lewis, D. (2007). David temperley, music and probability (review). *Empirical Musicology*, *2*(4), 155-163.

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2008). A comparison of statistical and rule-based models of melodic segmentation. In *Proceedings of the 9th international conference on music information retrieval.*

Pearce, M. T., & Wiggins, G. A. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, *33*(4), 367-385.

Pearce, M. T., & Wiggins, G. A. (2006, June). Expectation in melody: The influence of context and learning. *Music Perception*, *23*(5), 377-405.

Peper, C., Beek, P., & Wieringen, P. van. (1995). Multifrequency coordination in bimanual tapping: Asymmetrical coupling and signs of supercriticality. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 117-1138.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis.* Berlin: Springer.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in schumann's träumerei. *Journal of the Acoustical*

Society of America, *92*(5), 2546-2568.
Rhodes, C., Lewis, D., & Müllensiefen, D. (2007). Bayesian model selection for harmonic labelling. In *Program and summaries of the first internatinal confernece of the society for mathematics and computation in music, berlin, may, 18-20, 2007.*
Sadakata, M., Desain, P., & Honing, H. (2006, February). The bayesian way to relate rhythm perception and production. *Music Perception*, *23*(3), 269-288.
Sahmer, K., Vigneau, E., & Qannari, E. (2006). A cluster approach to analyze preference data: Choice of the number of clusters. *Food Quality and Preference*, *17*(3-4), 257-265.
Schellenberg, E. G. (1997). Simplifying the implication-realization model of simplifying the implication-realization model of melodic expectancy. *Music Perception*, *14*, 295-318.
Schubert, E., & Dunsmuir, W. (1999). Regression modelling continuous data in music psychology. In S. W. Yi (Ed.), *Music, mind, and science* (p. 298-352). Seoul: National University Press.
Schumaker, L. (1981). *Spline functions: Basic theory.* John Wiley and Sons.
Steinbeck, W. (1982). *Struktur und ähnlichkeit: Methoden automatisierter melodieanalyse*. Kassel: Bärenreiter.
Temperley, D. (2001). *The cognition of basic musical structures.* Cambridge, MA: MIT Press.
Temperley, D. (2004, Fall). Bayesian models of musical structure and cognition. *Musicae Scientiae*, *8*(2), 175-205.
Temperley, D. (2006). A probabilistic model of melody perception. In *Proceeding of the 7th international conference on music information retrieval.*
Temperley, D. (2007). *Music and probability.* Cambridge, MA: MIT Press.
Therneau, T., & Atkinson, E. (1997). *An introduction to recursive partitioning using the rpart routines* (Tech. Rep.). Mayo Clinic Division of Biostatistics.
Tillmann, B., Bharucha, J. J., & Bigand, E. (2003). Learning and perceiving musical structures: further insights from artificial neural networks. In R. Zatorre & I. Peretz (Eds.), *The cognitive neuroscience of music* (p. 109-123). Oxford: Oxford University Press.
Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*, 80-113.
Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005, December). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques. *Music Perception*, *23*(2), 137-152.
Windsor, L. (2001). Data collection, experimental design, and statistics in musical research. In E. Clarke & N. Cook (Eds.), *Empirical musicology: Aims,*

methods, prospects (p. 197-222). Oxford: Oxford University Press.
Winsberg, S., & De Soete, G. (1993). A latent class approach to fitting the weighted euclidean model, clascal. *Psychometrika*, *58*, 315-330.
Wundt, W. (1882). Über musikpsychologische methoden. In *Philosophische schriften* (Vol. 1, p. 138).