# *Classification of high and low achievers in a music sight-reading task*

Psychology of Music

*Psychology of Music*

vol 34(1): 5–26 [0305-7356 (200610) 34:1; 5–26]
10.1177/0305735606059102
http://pom.sagepub.com

REINHARD KOPIEZ
HANOVER UNIVERSITY OF MUSIC AND DRAMA, GERMANY
CLAUS WEIHS
UNIVERSITY OF DORTMUND, GERMANY
UWE LIGGES
UNIVERSITY OF DORTMUND, GERMANY
JI IN LEE
HANOVER UNIVERSITY OF MUSIC AND DRAMA, GERMANY

ABSTRACT The unrehearsed performance of music, called 'sight-reading' (SR), is a basic skill for all musicians. It is of particular interest for musical occupations such as the piano accompanist, the conductor, or the correpetiteur. However, up until now, there is no theory of SR which considers all relevant factors such as practice-related variables (e.g. expertise), speed of information processing (e.g. mental speed), or psychomotor speed (e.g. speed of trills). Despite the merits of expertise theory, there is no comprehensive model that can classify subjects into high- and low-performance groups. In contrast to previous studies, this study uses a data mining approach instead of regression analysis and tries to classify subjects into predetermined achievement classes. It is based on an extensive experiment in which the total SR performance of 52 piano students at a German music department was measured by use of an accompanying task. Additionally, subjects completed a set of psychological tests, such as tests of mental speed, reaction time, working memory, inner hearing, etc., which were found in earlier studies to be useful predictors of SR achievement. For the first time, classification methods (cluster analysis, regression trees, classification trees, linear discriminant analysis) were applied to determine combinations of variables for classification. Results of a linear discriminant analysis revealed a two-class solution with four predictors (cross-validated error: 15%) and a three-class solution with five predictors (cross-validated error: 33%).

KEYWORDS: *music performance, music reading, performance analysis*

Sight-reading (SR) is a skill which is required by all musicians. It is not only of particular interest for musical occupations such as the piano accompanist, the conductor, or the correpetiteur, but is also one of the five basic performance skills every musician should acquire. McPherson (1993) defines these skills as an ability to perform rehearsed repertoire, perform from memory (where music is memorized using notation and then re-created aurally), play by ear (where music is both learned and reproduced aurally), improvise and sight-read music without prior rehearsal. This last skill is characterized by high demands on the performer's capacity to process complex visual input (the score) under real-time constraints and without the opportunity of error correction. (For a more detailed explanation of the flow of influences between these skills, see McPherson, 1995; McPherson and Gabrielsson, 2002; McPherson et al., 1997.)

However, up until now, differences between individuals in SR achievement have not yet been fully understood. There has been no theory of SR that considers all relevant factors such as practice-related variables, speed of information processing, or psychomotor speed. From previous studies, we already know that there are a number of music-specific and non-music-specific skills which are relevant to the explanation of differences in SR performance. A study by Kornicke (1992, 1995), based on 73 piano students, revealed the following influential variables: (1) aural imagery; (2) SR experience (quantity, frequency, and range of SR); (3) cognitive style of field dependence/field independence (important for males); (4) style of thinking measured by the Myers-Briggs Type Indicator; (5) external locus of control (which is a tendency to perceive reinforcement as the result of external 'forces' such as chance or fate). Another influential study by Lehmann and Ericsson (1996), based on expertise theory, measured the performance of 16 expert piano accompanists and revealed the following variables as best predictors in a multiple regression analysis: (1) accumulated amount of time spent on accompanying-related activities; (2) size of accompanying repertoire. (For a detailed discussion of the relationship between subskills of SR see also Lehmann and McArthur, 2002).

In our study, considered subskill variables were selected from three groups: (1) general cognitive skills; (2) elementary cognitive skills; and (3) practice-related skills (see Table 1 later). For the group of elementary cognitive skills our selection of music-specific and non-music-specific memory skills was derived from the typical demands of the task. From Sloboda's (1974) early studies we know that the ability to read ahead while playing is a condition for successful SR. However, without a sufficient short-term memory buffer, the advantage of an extended eye–hand span remains useless. Surprisingly, there are only a few studies that investigate the influence of short-term memory on SR achievement (e.g. Eaton, 1978; Waters et al., 1998). Although, SR is a typical task with high demands on working memory (this kind of memory is characterized by the simultaneous storage and processing of information), up until now this kind of memory has not been considered in SR research. Thus, music-specific as well as non-music-specific tests for short-term and working memory have been included in our study. Additionally, based on the findings by Salis (1978), we also included aspects of general mental capacity (e.g. indicated by IQ) through use of a subsection of the Raven matrices (Raven, 2000). Salis found a correlation of $r$ = .57 for the total IQ score (measured by the Wechsler Adult Intelligence Scales) between SR performance and a correlation of $r$ = .39 for musical short-term memory (measured by the Drake Musical Aptitude Test).

For the group of elementary cognitive skills, selection of predictors was based on the assumption that under the time constraints of SR, the speed of information intake and of information processing plays a crucial role. For example, as Eaton (1978) could show, psychomotor skills (speed of key identification) is an important predictor of SR achievement. Thompson (1985) regarded SR as a transcription task and he assumed that speed of information processing and reaction time played a crucial role. He included a musical reaction time task and found a correlation of $r$ = -.54 between the number of correctly performed measures per second of the Watkins-Farnum Performance Scale and the time needed to perform a suddenly appearing note.

Thus, simple reaction time, psychomotor movement speed (wrist tapping and speed trill) have been considered. Additionally, laterality (handedness) has been included as an influential variable. This selection was based on the observation by Kilshaw and Annett (1983) that left-handers tend to be faster than right-handers in a peg-moving task. However, this effect was only clear for the non-preferred hand. Authors assume that the 'typical human bias to the left hemisphere and right hand might be due to a right hemisphere handicap rather than to a left hemisphere advantage' (Annett and Kilshaw, 1983: 269). This means that a weaker right-hand lateralization (and stronger tendency to non-right-handedness) could result in better SR performance. The first study that included aspects of visual field laterality in a tachistoscopic task and its influence on SR performance was conducted by Salis (1978). However, she could not find an influence of laterality, because all subjects except one were right-handed.

For the group of practice-related skills, of course, we have to consider acquired SR and accompanying expertise. The importance of accumulated practice in the domain of SR has been explained in the ground-breaking studies by Lehmann and Ericsson (1993, 1996). Auditory imagery (inner hearing) has been considered because the study by Schleuter (1993) revealed some correlation between audiation and SR achievement ($r$ = .25). Kornicke (1995) states that audiation is the highest predictor and Waters et al. (1998) confirmed the importance of audiation.

On this background, in a recent study, Lee (2004) used a set of more than 20 predictors derived from these groups of skills to determine those that best can explain differences in SR achievement by multiple regression analysis. Trill speed, accumulated SR expertise up to the age of 15, speed of information processing and speed of tapping were identified as the best predictors and could explain 65 percent of variance (adjusted $R^2$ = 0.65).

However, our study tries a new methodological approach to explain inter-individual differences in SR achievement and is based on the assumption that, given a limited number of subjects, a method of data analysis which searches for clusters that represent a sufficient number of cases and are separable with an acceptable error is more promising than regression methods, which search for sophisticated linear or non-linear relationships. In regression analysis, the attempt is made to predict a score from some explanatory variables. If there is only a limited number of subjects in the study, and variances are rather large, confidence intervals are very large as well. Therefore, significance of some predictors cannot be shown by statistical tests. Using classification methods, we try to separate just two or three classes of achievers. Hence, we predict two-class classification of cases such as – vs. +, or three-class classification such as –/0/+, respectively. Even in a study with a few subjects and rather large variances it might be possible to get a fairly good separation of the classes, while meaningful prediction of scores by means of linear regression is impossible. To achieve our aim, we applied classification methods to uncover those variables or variable combinations that best contribute to the classification of SR into performance classes.

The motivation for this study is two-fold: first, this approach applies methods of classification to SR performance for the first time. These statistical methods are well known in many applications, but have rarely been applied in the field of music psychology, even though some of the properties are very promising. Second, this study extends the expertise approach and tests the hypothesis that SR expertise (e.g. measured by the accumulated hours of domain-specifc practice) is a necessary, yet not entirely sufficient predictor of SR performance.

## *Method*

### SUBJECTS

Subjects were 52 piano students (28 females, 24 males) from the Hanover University of Music and Drama (mean age = 24.56, SD = 4.9). These pianists had to have piano as a major or be experts in piano chamber music or accompanying. Subjects were paid 30 euros for participation.

### MATERIALS

For the SR task, the paradigm of a pre-recorded pacing melody was used (Lehmann and Ericsson, 1993). This method creates time constraints that force the subjects to play in tempo. Materials consisted of two warm-up pieces and five pieces of increasing complexity. These were taken from existing piano SR literature (UNISA, n.d.), and a composer rearranged these pieces for a solo melody and piano accompaniment. The pre-recorded solo melody was played metronomically (i.e. synchronized to a metronome) by a violinist. Before each piece, tempo indications were given by clicks, which were also pre-recorded. Usually, these clicks were two full bars and this also gave the subjects an indication of when they should start playing.

### PROCEDURE

Subjects were required to accompany the pre-recorded violin part on a MIDI piano. The accompaniment was recorded to a PC by use of the sequencer software Cubase® (manufactured by Steinberg, Hamburg). Retrospective interviews and measurement of predictor variables were carried out after the SR tasks. The entire procedure lasted about 3 hours.

Scoring for the SR performances (target variable) was done using a researcher-developed computer program called MidiCompare (Dixon, 2002). This program matches the pitches of a subject's recorded SR performance with the score. The output shows the number of matches within an adjustable critical timeframe of ± 0.25 seconds. For this analysis, the total performance score of each subject for both hands as a percentage is used (matching pitches/total pitches in the score × 100).

### MEASUREMENT OF PREDICTOR VARIABLES

Selected predictor variables were derived from SR literature and divided into three groups: (1) general cognitive skills (such as short-term and working memory); (2) elementary cognitive and psychomotor skills (such as simple reaction time and speed of information processing); and (3) practice-related skills (such as general piano expertise, inner hearing ability and accumulated

TABLE 1 *Grouped predictor variables and methods used for measurement considered in the sight-reading (SR) experiment*

| Group of skills | Test | Method of measurement |
|---|---|---|
| 1. General cognitive skills | Non-music-specific short-term memory | Non-music-specific short-term memory test (Oberauer et al., 2000) |
| | Working memory | Working memory test (Oberauer et al., 2000) |
| | Music-specific short-term memory | Music-specific short-term memory test (Drösler, 1989) |
| | General intelligence | Raven's D matrices |
| 2. Elementary cognitive skills | Reaction time | Researcher-developed simple reaction time test (visual, auditory) |
| | Tapping speed | Wrist tapping |
| | Speed of information processing | Number connection test (Oswald and Roth, 1997) |
| | Psychomotor movement speed | Speed trilling for 15 seconds |
| | Lateralization | 'Objective' handedness by tapping performance |
| 3. Practice-related skills | Auditory imagery | Researcher-developed inner hearing test |
| | Sight-reading expertise | Retrospective interview for sight-reading and piano expertise |

TABLE 2 *List of 27 independent variables used for classification of sight-reading (SR) performance*

| Variable name | Variable label |
|---|---|
| ACHSRE10 | Accumulated hours of SR expertise up to age 10 |
| ACHSRE15 | Accumulated hours of SR expertise up to 15 |
| ACHSRE18 | Accumulated hours of SR expertise up to 18 |
| ACHSRETT | Accumulated hours of SR expertise total |
| ACHPSO10 | Accumulated hours of solo practice up to 10 |
| ACHPSO15 | Accumulated hours of solo practice up to 15 |
| ACHPSO18 | Accumulated hours of solo practice up to 18 |
| ACHPSOTT | Accumulated hours of solo practice total |
| ACYPLE10 | Accumulated hours of piano lessons up to 10 |
| ACYPLE15 | Accumulated hours of piano lessons up to 15 |
| ACYPLE18 | Accumulated hours of piano lessons up to 18 |
| ACYPLETT | Accumulated hours of piano lessons total |
| IH.DPRIM | Inner hearing score (d') |
| STMSMT | Short term music specific memory (no. of notes) |
| NUMCONTS | Number connection test (s) |
| RAVENMDS | Raven's D matrices (no. of correct items) |
| TOTTRAVE | Total time for Raven's D matrices (s) |
| STM.PER | Short term memory (mean% of correct items) |
| WM.PERC | Working memory (mean% of correct items) |
| PICRT.ME | Reaction time picture (median in ms) |
| SNDRT.ME | Reaction time sound (median in ms) |
| ITI.LRHZ | Inter tap interval for both hands (median in Hz) |
| TR131HZ* | Trill speed over 15 s, f.c.** 1-3, 1st trial (median in Hz) |
| TR132HZ* | Trill speed over 15 s, f.c.** 1-3, 2nd trial (median in Hz) |
| TR341HZ* | Trill speed over 15 s, f.c.** 3-4, 1st trial (median in Hz) |
| TR342HZ* | Trill speed over 15 s, f.c.** 3-4, 2nd trial (median in Hz) |
| LC | Tapping lateralization coefficient (< 1.7 = non-right-handed) |

* All trills were played with the right hand.
** f.c. = finger combination; all trills were played with the right hand.

hours of SR expertise) (for the grouping of skills see Table 1). In total, there were 27 single predictors considered (see Table 2). Although most of the predictors are well-known in SR literature, some unusual variables have been considered. Thus, the two variables 'lateralization' and 'inner hearing' are explained in more detail later.

### *Lateralization*

From Schlaug's studies (Schlaug et al., 1995; Schlaug, 2001) we know that due to neural plasticity, the brain of professional musicans is characterized by a stronger connection between the two hemispheres (corpus callosum). A stronger connection between hemispheres can also be interpreted as an improved communication between hemispheres and a weaker lateralization for one hand. To test this prediction, we had to determine the subjects' handedness.

The determination of handedness in this study is based on Annett's (1985) 'right-shift theory'. The two-component theory assumes that handedness in humans and non-human primates depends on chance but that change in handedness is weighted towards right-handedness in most people by an agent of right-hemisphere disadvantage. It argues for the existence of a single gene for right shift (RS+) that evolved in humans to aid the growth of speech in the left hemisphere of the brain. This means that only the preference for right-handedness is determined genetically, and whether non-right-handed people become left-handed or ambidextrous is a matter of chance and environmental influences. Annett distinguishes between a 'subjective' preference handedness (as measured for example by the Edinburgh Inventory, see Oldfield, 1971) and the 'objective' or 'true' performance handedness, which means the underlying genetically determined handedness. Tests for preference handedness allocate subjects to the two handedness categories 'right' and 'left'. Tests for performance handedness assume a continuous and right-shifted distribution of handedness in the general population. Objective handedness as an environment-free indicator for genetically determined handedness can only be measured by means of a performance task (e.g. speed tapping). This method was first introduced by Peters and Durding (1978). In our study we also used a speed tapping task over 30 seconds for both hands. Measurement was made with a morse key (model by Junker Ltd., Germany; trigger point = 300 grams), connected to a PC and recorded through researcher-developed software. Wrist tapping was used as the task (movement was controlled by wrist, the wrist remained on the desk, and fingers 2 and 3 were released from the key after each tap). The lateralization coefficient LC was calculated by the equation LC = 100 * (LCUM – RCUM)/(LCUM + RCUM) where LCUM and RCUM is the median of the intertap intervals of the left and right hand over a total tapping duration of 30 seconds. According to Annett's (1985) theory, the 'objective' (tapping) handedness in the form of a designated right-handedness and a designated non-right-handedness is determined by a threshold for the LC of 1.7. In a previous study (Stein, 1995), this value was calculated based on a sample of 441 students, and subjects with an LC > 1.7 were designated to be objective right-handers, whereas subjects with an LC < 1.7 were classified as non-right-handers. This threshold was computed as the mean of the overlap between the distribution of LC values of all 441 subjects in the two self-declared groups of right-handed vs. non-right-handed subjects.

*Inner hearing*

The term 'inner hearing' means the ability to imagine the sound of a score without the help of an instrument or singing. Based on the assumption that excellent SR is not only the result of an extended eye–hand span but of the ability to generate an aural image (audiation) of the printed score (Kornicke, 1992, 1995), we decided to consider the skill of inner hearing. The study by Waters et al. (1998) on predictors of SR performance showed that a highly developed skill for inner hearing could be of benefit due to a priming effect. However, the question remains open whether there is enough time to build up an aural image under the extreme time constraints of SR.

An early attempt to measure auditory imagery was made by Pagan (1970) using a simple chord imagery task. In our study, measurement of inner hearing was made according to the 'embedded melody paradigm' proposed by Brodsky et al. (2003). Subjects were given 45 seconds to look at a variation of a theme. In the next step, the original theme, or a 'lure melody', was heard through speakers and could be repeated. Subjects had to decide if the melody heard was embedded in the variation seen in the score (same–different paradigm). Stimuli lasted for less than 1 minute. We used a selection of themes and variations from classical piano literature resulting in a total of five inner hearing tasks from which a d-prime value (Macmillan and Creelman, 1991) for inner hearing was calculated.

## *Results*

The main aim of the classification analysis was to find variables that could classify cases with respect to the target variable 'total sight-reading performance' with a minimum error. Calculation of statistical analyses was done using the open source software R (R Development Core Team, 2004). The analysis comprised three steps:

1. We tried to find a solution that classified subjects into two classes.
2. We compared these findings to a solution with three classes
3. We applied the method of regression tree analysis to reveal the structure of classifiers.

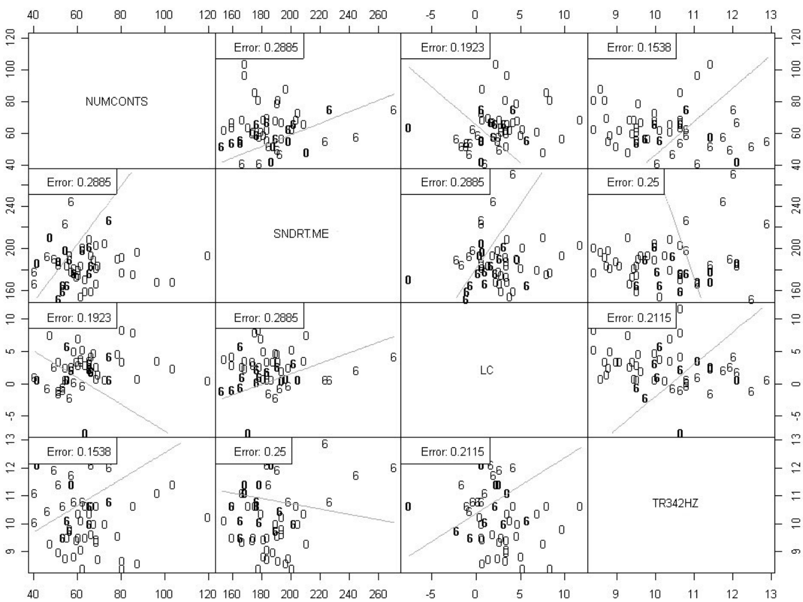
Finally, all solutions from the classification analysis were compared with the results obtained from the multiple regression analysis. Commonalities and differences between classification and regression analysis will be discussed below.

### THE TWO-CLASS SOLUTION

#### *Cluster analysis and 2-class LDA*

All 27 predictor variables were included (Table 2) and the total SR performance was used as the target variable. Analysis began with a separation of subjects into two total SR performance classes through a cluster analysis with method *k*-means (for a general description of classification procedures see Breiman et al., 1984; Venables and Ripley, 2002). Group boundaries were determined by the mean of the two cluster center values, resulting in two groups (0–66%, 66–100% performance). This separation of performance data into ranges of the lower two thirds and the upper one third, with group sizes of 33 and 19 subjects, is reasonable. Cases were classified on the basis of the 27 predictor variables using stepwise linear discriminant analysis (LDA; method: stepwise variable selection based on inner 4-fold cross-validation, direction: both, stop criterion: error improvement < 5%).

Four separating variables (LC, NUMCONTS, TR342HZ, SNDRT.ME) were revealed as classifying variables. Classification was successful with a total cross-validated error of 0.15 (= 15%; 4-fold cross-validated; see Table 3).[1] Despite the acceptable total error of 15 percent, we can see that particular variable combinations differ in error rate. Figure 1 shows a differentiated picture of the apparent error for each combination of two selected separating variables. Class boundaries are indicated by the grey classification line. Apparent error ranges between 0.154 (variable combination NUMCONTS-TR342HZ) and 0.288 (variable combinations NUMCONTS-SNDRT.ME and LC-SNDRT.ME). On the one hand, the combination of an elementary cognitive skill, such as simple reaction time in an auditory task (SNDRT.ME), with a psychomotor skill component (speed trill TR342HZ) reveals that a subject with a slower trill speed (< 11 Hz) and a shorter reaction time (< 200 ms) can be classified to the upper third performance class (66–100%) with an apparent error of 25 percent. On the other hand, a combination of right-handedness (LC > 2) and a relatively slow mental speed (NUMCONTS > 60 s) also classifies subjects to the upper third performance class.



*Note:* Grey line indicates class boundaries. Case-allocation to classes is indicated by symbol '0' for 0–66% and '6' for 66–100%. Bold digits indicate false classification of cases to the respective class. The apparent error for each combination of separating variables is indicated in the upper left corner of each box.

FIGURE 1 *Error matrix scatterplot for two-class linear discrimination ananlysis (LDA) (0, 66, 100) with total sight-reading (SR) achievement as the target variable.*

TABLE 3 *Cross-validated error rate of two- and three-class clasifications by LDA for total sight-reading (SR) achievement*

| Target variable | Classes *N* | Separating variables | Range of classes (%) (indicator) | Subjects *N* | Error rate (%) |
|---|---|---|---|---|---|
| Total achievement | 2 | 4 (LC, NUMCONTS, TR342HZ, SNDRT.ME) | 0–66 (0)<br>66–100 (6) | 33<br>19 | 15 |
| | 3 | 5 (ACHSRE15, TR132HZ, ACHSRE18, TR131HZ, TR342HZ) | 0–35 (0)<br>35–80 (3)<br>80–100 (8) | 25<br>15<br>12 | 33 |

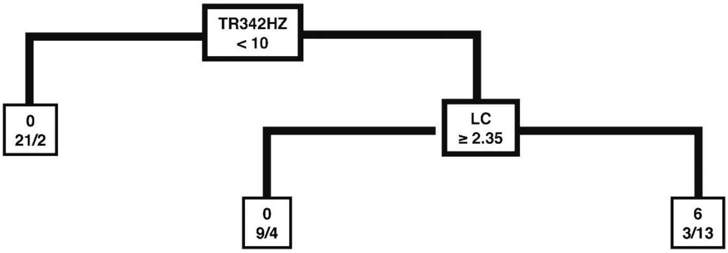
*Note*: Separating variables are listed in sequence of data entry into LDA (method: stepwise variable selection, both directions).

### *The 2-class classification tree*

A clearer but less differentiated picture of classifying variables is given by the 2-class classification tree (Figure 2) (for a statistical description of classification trees see Therneau and Atkinson, 1997). All 27 independent variables were included in the two-class tree analysis leading to a tree with a relatively high error of 0.44 (10-fold cross-validated). Figure 2 should be read as follows: the left side of the first branch allocates those subjects to the lower performance group 0 (0–66%), whose median trill speed over 15 seconds with the third and fourth finger is slower than 10 Hz. This left branch comprises 23 cases. Twenty-one cases could be classified correctly to class 0 and two subjects from class 6 (66–100%) were misclassified to class 0. Those subjects who could trill faster than 10 Hz ($n$ = 29) were sorted into the right branch of the top-level node. In the case of a tendency to non-right-handedness (the LC value should be smaller than about 1.7), these subjects were sorted on the next deeper branch level into the right branch and high-performance class 6 (66–100%). In this lower right branch of the classification tree, 13 out of 16 subjects could be classified correctly to the 66–100% group and three cases were misclassified using these two criteria. In the case of a trill speed greater than 10 Hz and a tendency to right-handedness, nine subjects were classified correctly to the performance class 0 (0–66%; left branch of node LC) and four cases were misclassified.

## THE THREE-CLASS SOLUTION

In this classfication step, we tried to improve our two-class solution by adding a third performance class. It was assumed that the three-class procedure would provide a more detailed insight into classificator combinations compared to the two-class solution. The following classification procedure was used:




*Note*: Left side of the branch means 'yes' for the classification condition and right side means 'no'.

FIGURE 2 *Classification tree for the two-class solution (0, 66, 100 = 0–66% and 66–100%) with an error of 0.44 (10-fold cross-validated).*

1. calculation of the three class ranges by means of a cluster analysis;
2. calculation of a three-class LDA;
3. calculation of a three-class classification tree;
4. calculation of a regression tree.

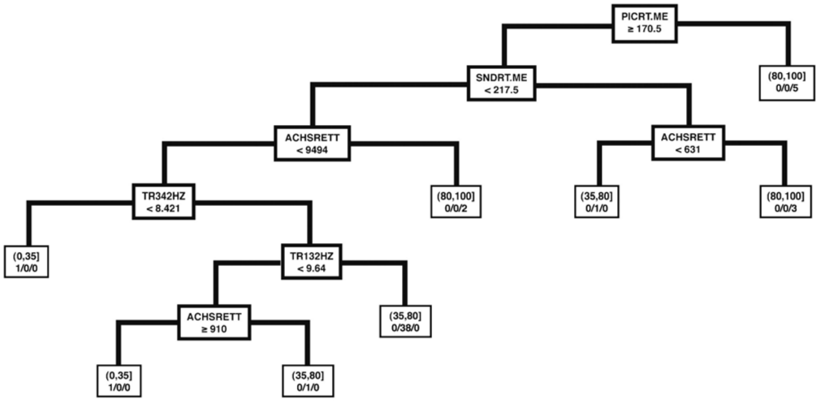
*Cluster analysis and three-class LDA*

To determine class ranges for the classification procedure, a cluster analysis (method: *k*-means) was conducted in the next step, and class ranges (0, 35, 80, 100) were obtained by averaging cluster center values with group sizes of $n$ = 25, 15 and 12 subjects (see Table 3). Class ranges were used for the resultant LDA (method: stepwise variable selection based on leave-one-out cross-validation, direction: both, stop criterion: error improvement < 1%). The three-class LDA resulted in a higher cross-validated error of 0.33. In some variable combinations, apparent error for classification exceeds 0.40. Thus, selected variable combinations will not be discussed in detail.

*The three-class classification tree*

All 27 independent variables were entered into the three-class tree analysis. Classification was successful, with an error of 0.19 (10-fold cross-validated). Compared to the two-class classification tree, the three-class classification tree results in a much better classification of cases in the three predefined classes (breaks at 0, 35, 80, 100). The three-class classification (see Figure 3) originates from an elementary cognitive skill (simple reaction time for a visual stimulus; PICRT.ME) and reveals three possible ways of reaching the high-performance class (80–100%):

1. Five cases had a reaction time shorter than 170.5 ms and could be classified in the top node (PICRT.ME) (80–100%).
2. The left branch of the top node PICRT.ME is characterized by subjects with a slow simple reaction time for a visual stimulus. If subjects also showed a slower reaction time for an auditory stimulus (right branch of node SNDRT.ME) and their accumulated total SR expertise (ACHSRETT)




*Note*: Left side of the branch means 'yes' for the classification condition and right side means 'no'.

FIGURE 3 *Classification tree for the three-class solution (0, 35, 80, 100 = 0–35%, 35–80% and 80–100%) with a total error of 0.19 (10-fold cross-validated).*

exceeded more than 631 hours, three cases could be sorted into the high-performance group.

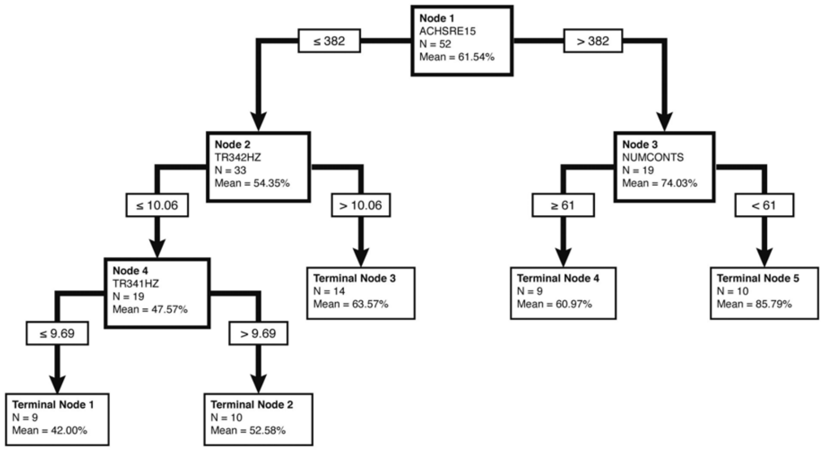
3. If subjects showed a fast reaction time for an auditory stimulus (left branch of node SNDRT.ME) and had accumulated more than 9494 hours of SR expertise (right branch of node ACHSRETT), two cases could also be classified into the high-performance group.

However, three-class classification reaches its limit with a total number of 52 subjects. Some terminal nodes are only represented by one to two subjects. The majority of subjects are sorted into the 35–80 percent group ($n$ = 40) and different combinations of the predictors simple reaction time, psychomotor speed and SR expertise result in this medium performance classification. Due to a total number of $n$ = 2, classification into the lowest performance class has to be interpreted with care.

To summarize the results of this complex tree analysis, we can first conclude that speed of reaction time and SR expertise are important in reaching a high-performance class, and there are simple as well as complex variable combinations, which are relevant for the classification into a medium and high class. Second, the classification tree reveals highly individual pathways into the high-performance group. From this result we might conclude that SR performance is not the result of a single predictor, such as expertise, but of a complex interaction between predictors.

*Regression tree*

As a last classification analysis step, regression tree analysis was done using the procedure 'rpart' from the software package R (R Development Core Team, 2004). The rpart algorithm (Therneau and Atkinson, 1997) is based on a similar algorithm used in the software CART® (Breiman et al., 1984;




*Note*: Average sight reading achievement for each node as a percentage and condition for node split is indicated. Left side of a branch means 'yes' for the node condition and right side means 'no'. For meaning of variable names see Table 2.

FIGURE 4 *Regression tree and predictor variables of total sight-reading (SR) performance obtained from tree analysis procedure 'rpart'. Resubstitution error: 0.33 (10-fold cross-validated). Minimal node size:* n = *15, minimal terminal node size:* n = *5.*

Crawford, 1987; Steinberg and Colla, 1997). To avoid an overfitting problem, adjustment of the rpart procedure was set to restrictive conditions (method: stepwise with 10-fold cross-validation; minimum node size: $n$ = 15, minimum terminal node size: $n$ = 5). As can be seen in Figure 4, the regression tree analysis using rpart results in a set of four variables (ACHSRE15, TR342HZ, NUMCONTS, TR341HZ), which constitute a five-terminal-node tree (apparent relative error: 0.33). Split criteria and statistical features of each node are indicated (left branch of a node means 'yes' for the split criteria and right branch means 'no'). Table 4 shows the descriptive statistics for each classifying node variable. The regression tree shows a differentiated picture of the classification analysis: starting from node 1, the right branch ends in two terminal nodes (4 and 5) with terminal node 5 representing the high-performance group with 85.79 percent of pitches correctly performed. Conversely, the left side of the node 1 branch ends in two terminal nodes (1 and 2) with less than 53 percent of total achievement. To understand the meaning of the nodes, the way from node 1 to terminal node 5 will be traced from top to bottom of the regression tree. The main separation in node 1 at the top level is made by an expertise-related variable (ACHSRE15; accumulated amount of SR expertise up to the age of 15). Subjects with more than 382 hours of SR expertise are allocated to the right branch.

Starting again from node 3, to achieve a better result, a subject has to be quick in the mental speed test (NUMCONTS < 61 s). This means that an average achievement of 86 percent is the result of a combination of two different variables (and two nodes) of a regression tree. On the other side, the

TABLE 4 *Descriptive statistics of classifications as revealed by regression tree analysis* (n = 52)

| | ACHSRE 15 (h) | ACHSRE 18 (h) | ACHSRETT (h) | NCT (s) | RT picture (median in ms) | RT sound (median in ms) | ITI both hands (median in Hz) | LC (< 1.7 = non-right-handed) | Trill 1–3–1 (median in Hz) | Trill 1–3–2 (median in Hz) | Trill 3–4–1 (median in Hz) | Trill 3–4–2 (median in Hz) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 454.46 | 948.42 | 2508.96 | 64.73 | 196.98 | 187.33 | 6.38 | 2.63 | 11.96 | 11.50 | 10.53 | 10.25 |
| Median | 212.00 | 572.00 | 1118.00 | 62.50 | 190.50 | 184.00 | 6.38 | 2.55 | 11.85 | 11.42 | 10.42 | 10.12 |
| SD | 668.05 | 1186.84 | 3184.72 | 15.09 | 21.97 | 21.72 | .55 | 3.21 | 1.07 | .98 | 1.00 | 1.11 |
| Minimum | 0 | 0 | 15 | 41 | 168 | 153 | 4.90 | –7.6 | 9.75 | 9.20 | 8.60 | 8.42 |
| Maximum | 3520 | 6528 | 13962 | 120 | 250 | 271 | 7.63 | 11.8 | 15.68 | 13.56 | 12.50 | 12.90 |

TABLE 5 *Multiple regression analysis of predictor variables and dependent variable 'total sight reading achievement' (method = stepwise)*

| Model | Variables | Adjusted $R^2$ | $\Delta R^2$ | Beta coefficient | $\rho$ |
|---|---|---|---|---|---|
| 1 | TR342HZ | .309 | | 0.568 | .000 |
| 2 | TR342HZ | .479 | .170 | .518 | .000 |
| | ACHSRE15 | | | .424 | |
| 3 | TR342HZ | 0.561 | .082 | .486 | .003 |
| | ACHSRE15 | | | .399 | |
| | TR341HZ | | | –.298 | |
| 4 | TR342HZ | .595 | .034 | .404 | .029 |
| | ACHSRE15 | | | .390 | |
| | TR341HZ | | | –.276 | |
| | NUMCONTS | | | .219 | |
| 5 | TR342HZ | .623 | .028 | .438 | .038 |
| | ACHSRE15 | | | .427 | |
| | TR131HZ | | | –.305 | |
| | NUMCONTS | | | .257 | |
| | ITI.LRHZ | | | –.202 | |
| 6 | TR342HZ | .651 | .028 | .570 | .038 |
| | ACHSRE15 | | | .410 | |
| | TR131HZ | | | –.325 | |
| | NUMCONTS | | | .500 | |
| | ITI.LRHZ | | | –.242 | |
| | TR132HZ | | | –.357 | |

*Note*: For explanation of variable names see Table 2.

way to the low-performance group (terminal node 1) leads from node 1 to node 2 in the case of less than 382 hours of accumulated SR expertise up until the age of 15. Node 2 splits up again into the left branch if trilling speed for the finger combination 3–4 (2nd trial) is slower than 10.06 Hz, and results into terminal node 1 if trilling speed for the finger combination 3–4 (first trial) is slower than 9.69 Hz. However, although the way to the high- and low-performance group is clear, the regression tree also reveals that there is more than one way to the mean range of performance (in the vicinity of about 60% of matched notes): low expertise and high trilling speed with combination 3–4–2 leads to terminal node 3 (63.57%) and higher expertise combined with slow mental speed leads to terminal node 4 with a comparable level of achievement (60.97%). In total, the relative resubstitution error of 0.33 with 33 percent of unexplained variance seems to show an acceptable match between the tree model and the data set.

*Commonalities and differences between classifiers*

Different methods of classification analysis revealed different classifications.

TABLE 6 *Commonalities of predictors in multiple regression analysis (MR), linear discriminant analysis with two and three classes (LDA2, LDA3), classification tree with two and three classes (2CT, 3CT), and regression tree (RT)*

| | TR342 | ACHSRE15 | TR132 | NUMCONTS | TR131 | LC | ACHSRE 18 | SNDRT. ME | PICRT. ME | ACHSRE TT | ITI | TR341 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MR | **×** | **×** | **×** | **×** | **×** | | | | | | × | |
| LDA 2 | **×** | | | **×** | | **×** | | × | | | | |
| LDA 3 | **×** | **×** | **×** | | **×** | | × | | | | | |
| 2CT | **×** | | | | | **×** | | | | | | |
| 3CT | **×** | **×** | **×** | | | | | | × | × | | |
| RT | **×** | **×** | | **×** | | | | | | | | × |

*Note*: For explanation of variable names see Table 2. Predictors with at least two commonalities are highlighted bold.

It must be borne in mind that the analysis by multiple regression resulted in an adjusted $R^2$ of 0.65 (see Table 5). However, this means that 35 percent of variance remained unexplained by using statistical methods of linear combinations of predictors. This was our initial motivation to look for an alternative method of performance prediction by use of classification procedures. As Table 5 shows, the optimal combination of predictor variables in multiple regression is reached in model 6, which includes trill speed, SR expertise up to the age of 15, speed of information processing (NCT) and wrist tapping speed. Compared with the results obtained from non-linear classification analysis (see Table 6), we can clearly see that most of the linear predictors are also identified as classifiers – except tapping speed (ITI). However, there is one important difference: hand lateralization (this means the tendency to right-handedness or non-right-handedness) remains unrevealed by multiple regression analysis but is revealed as a classifier by two-class linear discriminant analysis and two-class classification tree. The influence of hand lateralization on SR performance will be the subject of future data analyses.

## *Discussion*

We could demonstrate that classification of SR performance is a useful method of SR achievement analysis and results in an acceptable error. In the case of the two-class LDA we found a cross-validated misclassification error of 0.15. This means that just eight out of the 52 subjects were misclassified by cross-validation. This error rate is small enough to be acceptable and can be interpreted much better than predicted score values based on a regression analysis with $R^2 = 0.65$ (35% unexplained variance). The two-class classification tree emphasizes the subject's psychomotor speed and handedness. Subjects with high trilling speed and a tendency to non-right-handedness are classified into the high-performance group. However, this simple classification results in a high error of 0.44.

A greater differentiation of classificators is given by the three-class classification tree (Figure 3), and we can observe three different classification pathways: subjects with a very fast visual reaction time are likely to be directly classified to the 80–100 percent class. However, there are two alternative paths to the high-performance class: first, in the case of a slow visual reaction time, subjects can be classified as high performers if their auditory reaction time is *slow* but the total accumulated SR expertise is more than 631 hours; second, in the case of slow reaction time to visual stimuli, subjects can be classified as high performers if their simple reaction time to auditory stimuli is *fast,* but the total accumulated SR expertise is more than 9494 hours. Under these circumstances, extremely high SR expertise can compensate for a slow visual reaction time. This observation of a compensating function of variables is a new result in SR research. Although there is currently no theoretical model that can explain this basic function of speed of visual reaction time in SR, speed of information processing in an elementary cognitive task (e.g. simple reaction time) seems to play a crucial role. However, three-class classification leaves the problem open that some final classes only represent a small number of cases.

Another new finding of our study is the relevance of basic motor processes such as psychomotor speed (as shown in Table 3) measured by trilling speed. There is currently no theory that explains the connection between these basic and higher cognitive processes. Our hypothesis is that the transfer of fast psychomotor control-processes (e.g. trilling speed, reaction time) to subcortical basic systems relieves the cortical cognitive system of workload. In this way, new capacity for information intake in SR is made possible. In other words, SR performance is not only the result of cortical control but is also determined by the degree of integration of the spinal chord into basic psychomotor control processes.

A new aspect of classification analysis is added by the regression tree analysis. The regression tree emphasizes the role of domain-specific expertise as a predictor for SR performance. However, expertise must be acquired ideally before the age of 15. This is also one of the main findings of Lee (2004). A second main aspect is the central role of factors that can be labelled as 'mental speed' and 'psychomotor speed': time needed for the number connection test (NUMCONTS) (see Oswald and Roth, 1997) builds node 3, which directly leads into the high-performance group if mental speed is high. Figure 4 also reveals that there are two ways to the medium performance group (right branch from node 2 and left branch from node 3) but only one way to the high performance group (right branch of node 3).

How can these findings be interpreted under consideration of commonalities and differences? As revealed by Table 6, all five classification methods have some predictors in common with those obtained from multiple regression analysis. The higher the number of variables in common, the higher the validity of predictors found. However, variables not considered in classification procedures should not be interpreted as irrelevant for the explanation of differences in SR performance, even though they are not useful as classifiers. Trilling speed with both finger combinations 3–4 and 1–3, the time needed for the number connection test (mental speed indicator) and domain-specific expertise before the age of 15 are important classifiers. Thus, in a first rough approach we might conclude: 'speed matters'.

The tendency to non-right-handedness is the sixth important classifier. This last aspect, the influence of handedness on SR achievement, opens a completely new and promising insight into possible neuropsychological foundations of SR performance. As the two-class classification (Figure 2) shows, allocation of cases to the upper third achievement group (66–100%) was successful if subjects showed a high trilling speed and a tendency to mixed-handedness. The influence of handedness on SR achievement will be analysed thoroughly in a forthcoming study.

From the perspective of music education, results shed a new light on the relationship between practice-dependent skills and invariant abilities: we can conclude that SR achievement, at a very high level in expert pianists, is determined by acquired expertise as well as by invariant factors such as ‘speed of information processing’ and ‘psychomotor speed’.

## REFERENCES


Annett, M. (1985) *Left, Right, Hand and Brain: The Right Shift Theory*. Hillsdale, NJ: Lawrence Erlbaum.

Annett, M. and Kilshaw, D. (1983) ‘Right- and Left-Hand Skill: II. Estimating the Parameters of the Distribution of L–R Differences in Males and Females, *British Journal of Psychology* 74(2): 269–83.

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

Brodsky, W., Henik, A., Rubinstein, B.-S. and Zorman, M. (2003) ‘Auditory Imagery from Musical Notation in Expert Musicians’, *Perception & Psychophysics* 65(4): 602–12.

Crawford, S.L. (1987) *Resampling Strategies for Recursive Partitioning Classification with the CART Algorithm*. Ann Arbor, MI: University Microfilms International.

Dixon, S. (2002) MidiCompare (computer software). Vienna: Austrian Institute for Artificial Intelligence.

Drösler, A. (1989) ‘Visuelle Wahrnehmungen von Notenfolgen: eine experimentelle Untersuchung zum Kurzzeitgedächtnis von Experten- und Laienmusikern verschiedener Altersgruppen’ (Visual perception of note sequences: An experimental investigation of the short-term memory of expert and non-expert musicians of different age groups), PhD thesis, Technical University of Berlin, Germany.

Eaton, J.L. (1978) ‘A Correlation Study of Keyboard Sight-Reading Facility with Previous Training, Note-Reading, Psychomotor, and Memorization Skills’, Doctor of Musical Arts thesis, Indiana University (Dissertation Abstracts International, A 39/07, p. 4109).

Hastie, T.J., Tibshirani, R.J. and Friedman, J. (2001) *The Elements of Statistical Learning: Data Mining Inference and Prediction*. New York: Springer.

Kilshaw, D. and Annett, M. (1983) ‘Right- and Left-Hand Skill: I. Effects of Age, Sex and Hand Preference Showing Superior Skill in Left-Handers’, *British Journal of Psychology* 74(2): 253–68.

Kornicke, L.E. (1992) ‘An Exploratory Study of Individual Difference Variables in Piano Sight-Reading Achievement’, PhD thesis, Indiana University (Dissertation Abstracts International A 53/12, p. 4125).

Kornicke, L.E. (1995) ‘An Exploratory Study of Individual Difference Variables in Piano Sight-Reading Achievement’, *Quarterly Journal of Music Teaching and Learning* 6(1): 56–79.

Lee, J.I. (2004) ‘Component Skills Involved in Sight Reading Music’, PhD thesis, Hanover University of Music and Drama, Hanover, Germany.

Lehmann, A.C. and Ericsson, K.A. (1993) ‘Sight-reading Ability of Expert Pianists in the Context of Piano Accompanying’, *Psychomusicology* 12(2): 182–95.

Lehmann, A.C. and Ericsson, K.A. (1996) ‘Performance without Preparation: Structure and Acquisition of Expert Sight-reading and Accompanying Performance’, *Psychomusicology* 15(1–2): 1–29.

Lehmann, A.C. and McArthur, V. (2002) 'Sight-reading', in R. Parncutt and G.E. McPherson (eds) *The Science and Psychology of Musical Performance: Creative Strategies for Music Teaching and Learning,* pp. 135–50. Oxford: Oxford University Press.

Macmillan, N.A. and Creelman, C.D. (1991) *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.

McPherson, G.E. (1993) 'Factors and Abilities Influencing the Development of Visual, Aural and Creative Performance Skills in Music and their Educational Implications', PhD thesis, University of Sydney, Australia (Dissertation Abstracts International A 54/04, p. 1277).

McPherson, G.E. (1995) 'The Assessment of Musical Performance: Development and Validation of Five New Measures', *Psychology of Music* 23(2): 142–61.

McPherson, G.E. and Gabrielsson, A. (2002) 'From Sound to Sign', in R. Parncutt and G.E. McPherson (eds) *The Science and Psychology of Musical Performance: Creative Strategies for Music Teaching and Learning,* pp. 99–115. Oxford: Oxford University Press.

McPherson, G.E., Bailey, M. and Sinclair, K. (1997) 'Path Analysis of a Model to Describe the Relationship Among Five Types of Musical Performance', *Journal of Research in Music Education* 45(1): 103–29.

Oberauer, K., Suess, H.-M., Schulze, R., Wilhelm, O. and Wittmann, W.W. (2000) 'Working Memory Capacity – Facets of a Cognitive Ability Construct, *Personality & Individual Differences* 29(6): 1017–45.

Oldfield, R.C. (1971) 'The Assessment and Analysis of Handedness: The Edingburgh Inventory', *Neuropsychologia* 9(1): 97–113.

Oswald, W.D. and Roth, E. (1997) *Der Zahlen-Verbindungs-Test* (The number combination test). Göttingen: Hogrefe.

Pagan, K.R. (1970) 'An Experiment in the Measurement of Certain Aspects of Score Reading Ability, Doctor of Music Education thesis, Indiana University (Dissertation Abstracts International A 31/01, p. 1313).

Peters, M. and Durding, B.M. (1978) 'Handedness Measured by Finger Tapping: A Continuous Variable', *Canadian Journal of Psychology* 32(4): 257–61.

R Development Core Team (2004) *R: A Language and Environment for Statistical Computing* (computer software for Statistical Analysis, for description see http://cran.r-project.org). Vienna: R Foundation for Statistical Computing.

Raven, J.C. (2000) *Standard Progressive Matrices*. Florence: Organizazzioni Speciali.

Salis, D.L. (1978) 'The Identification and Assessment of Cognitive Variables Associated with Reading of Advanced Music at the Piano', PhD thesis, University of Pittsburgh (Dissertation Abstracts International A 38/12, p. 7239–40).

Schlaug, G. (2001) 'The Brain of Musicians: A Model for Functional and Structural Adaptation', in R.J. Zatorre and I. Peretz (eds) *The Biological Foundations of Music* (Annals of the New York Academy of Sciences, Vol. 930, pp. 281–99). New York: The New York Academy of Sciences.

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J.F. and Steinmetz, H. (1995) 'Increased Corpus Callosum Size in Musicians', *Neuropsychologia* 33(8): 1047–1055.

Schleuter, S.L. (1993) 'The Relationship of AMMA Scores to Sight Singing, Dictation, and SAT Scores of University Music Majors', *Contributions to Music Education* 20: 57–63.

Sloboda, J.A. (1974) 'The Eye–Hand Span: An Approach to the Study of Sight Reading, *Psychology of Music* 2(2): 4–10.

Stein, C. v. d. (1995) 'Gibt es eine objektive Händigkeit? (Is there an objective handedness?, Diploma thesis, Department of Psychology, University of Cologne, Germany.

Steinberg, D. and Colla, P. (1997) *CART-Classification and Regression Trees, V 4.0* (computer software). San Diego, CA: Salford Systems.

Therneau, T.M. and Atkinson, E.J. (1997) 'An Introduction to Recursive Partitioning Using RPART Routines' (technical report), Mayo Foundation, Minnesota.

Thompson, W.B. (1985) ' Sources of Individual Differences in Music Sight-Reading Skill', PhD thesis, University of Missouri – Columbia (Dissertation Abstracts International B 47/02, p. 828).

UNISA (n.d.) *Playing at Sight (Piano)* (Vol. 1–8), Pretoria: University of South Africa.

Venables, W.N. and Ripley, B.D. (2002) *Modern Applied Statistics with S* (4th edition). New York: Springer.

Waters, A.J., Townsend, E. and Underwood, G. (1998) 'Expertise in Musical Sight-reading: A Study of Pianists', *British Journal of Psychology* 89(1): 123–149.

## NOTES

1. 'Cross-validation' is used to estimate the prediction error when a learned rule is applied to a new, independently drawn (test-)sample of observations. The data is divided into a certain number ($K$) of roughly equal-sized parts (e. g. $K = 4$, which accordingly is called a 4-fold cross-validation). For the prediction of the k-th out of $K$ parts, the model is fitted to the other $K$-1 parts and the error is calculated by predicting the $k$-th part, comparing the results of prediction with the known values. The averaged error from all $K$ parts is called 'cross-validated' error (for statistical details see Hastie et al., 2001; Venables and Ripley, 2002). In contrast, the 'apparent error' is the error we get when fitting the model to all observations and predicting all observations from this model again.

REINHARD KOPIEZ holds a degree in classical guitar, and Masters and PhD degrees in musicology. From 2000–03 he was vice-president of ESCOM and organizer of the 2003 5th Triennial ESCOM (European Society for the Cognitive Sciences of Music) Conference in Hanover, Germany. Since 2001 he has been president of the German Society for Music Psychology (DGM). He holds a position as Professor of Music Psychology at the Hanover University of Music and Drama. His main specialist area is performance research with a special interest in intonation, long-term performance (Satie's 'Vexation') and sight-reading.
*Address*: Hanover University of Music and Drama, Emmichplatz 1, 30175 Hanover, Germany. [email: kopiez@hmt-hanover.de]

CLAUS WEIHS holds a diploma and a PhD in mathematics. He was made Professor of Statistics in 1995 and holds the chair for Computer Aided Statistics at the University of Dortmund. Since 1997, he has been Deputy Director of the collaborative research center 'Reduction of Complexity for Multivariate Data Structures' (SFB 475) and is also President of the 'Gesellschaft für Klassifikation' (GfKl, German Society for Classification). His principal research areas are numerical methods, expert systems for explorative data analysis and design of experiments, statistical methods in quality control, music and statistics as well as classification methods.
*Address*: University of Dortmund, Department of statistics, Vogeslpothsweg 87, 44221 Dortmund, Germany. [email: weihs@statistik.uni-dortmund.de]

UWE LIGGES holds a diploma in statistics. Since 2001 he has been a faculty member of the Department of Statistics and a member of the collaborative research center 'Reduction of Complexity for Multivariate Data Structures' (SFB 475) at the University of Dortmund. His areas of interest are music and statistics as well as the statistical programming language R.
*Address*: University of Dortmund, Department of statistics, Vogeslpothsweg 87, 44221 Dortmund, Germany. [email: ligges@statistik.uni-dortmund.de]

JI IN LEE started playing the piano at the age of 6 and completed her BMus, B Mus (Hon) and MMus in South Africa in piano and in pedagogy. She has just completed her PhD in Music Psychology at the Hanover University of Music and Drama with a study of the sight-reading of music.
*Address*: Hanover University of Music and Drama, Emmichplatz 1, 30175 Hanover, Germany. [email: lee_ji_in1976@yahoo.co.uk]